

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 21, 2001, 10:47:22 ; Search time 43.71 Seconds
(without alignments)
5820.960 Million cell updates/sec

Title: US-09-596-958-1
Perfect score: 1344
Sequence: 1 atgtcaattcttagcttaa.....acctgaagggtgctgaatga 1344

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 324599 seqs, 94655562 residues

Total number of hits satisfying chosen parameters: 649198

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents_NA: *
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq: *
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq: *
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq: *
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq: *
5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq: *
6: /cgn2_6/ptodata/2/ina/backfiles1.seq: *

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	155.2	11.5	1729	4	US-09-120-817-1 Sequence 1, Appl
2	94.8	7.1	591	4	US-09-402-668-1 Sequence 1, Appl
3	74.2	5.5	666	4	US-09-198-956-3 Sequence 3, Appl
4	58.6	4.4	185	4	US-09-402-668-9 Sequence 9, Appl
5	45.8	3.4	7218	1	US-08-232-463-14 Sequence 14, Appl
6	43.2	3.2	7218	1	US-08-232-463-14 Sequence 14, Appl
7	37.4	2.8	2371	2	US-08-343-443B-1 Sequence 1, Appl
8	36.2	2.7	400	3	US-08-961-083-187 Sequence 187, App
9	35.8	2.7	53526	3	US-08-658-136-2 Sequence 2, Appl
10	35.8	2.7	53577	3	US-08-658-136-2 Sequence 1, Appl
11	35.2	2.6	1341	2	US-08-945-848-7 Sequence 7, Appl
12	35.2	2.6	2363	2	US-08-945-848-7 Sequence 6, Appl
13	34.6	2.6	1931	2	US-09-130-114-2 Sequence 2, Appl
14	34.2	2.5	485	5	PCT-US95-05741-10 Sequence 10, Appl
15	34	2.5	456	1	US-08-233-788A-58 Sequence 58, Appl
16	33.2	2.5	1849	2	US-08-676-166A-1 Sequence 1, Appl
17	32.8	2.4	1833	2	US-08-403-852D-6 Sequence 6, Appl
18	32.8	2.4	1833	3	US-08-510-646B-6 Sequence 6, Appl
19	32.8	2.4	1833	4	US-09-231-818-6 Sequence 6, Appl
20	32.4	2.4	3211	2	US-08-574-959A-8 Sequence 8, Appl
21	32.4	2.4	3901	2	US-08-574-959A-6 Sequence 6, Appl
22	32.2	2.4	1709	2	US-09-010-398-2 Sequence 2, Appl
23	32.2	2.4	1709	4	US-09-366-260-2 Sequence 2, Appl
24	32.2	2.4	1709	4	US-09-153-804-11 Sequence 11, Appl
25	32	2.4	913	1	US-08-217-327-3 Sequence 3, Appl
26	32	2.4	913	1	US-07-885-970A-3 Sequence 3, Appl
27	32	2.4	913	1	US-08-298-687A-3 Sequence 3, Appl

28	32	2.4	913	1	US-08-530-797-2	Sequence 2, Appl
29	32	2.4	913	2	US-08-298-829-3	Sequence 3, Appl
30	32	2.4	913	1	US-08-787-335-2	Sequence 2, Appl
31	32	2.4	1984	1	US-07-885-970A-25	Sequence 25, Appl
32	32	2.4	1985	1	US-08-298-687A-25	Sequence 25, Appl
33	32	2.4	1985	1	US-08-298-829-25	Sequence 25, Appl
34	32	2.4	3708	1	US-08-185-232A-1	Sequence 1, Appl
35	32	2.4	3708	1	US-08-416-523-1	Sequence 1, Appl
36	32	2.4	3708	3	US-08-789-478-1	Sequence 1, Appl
37	32	2.4	4481	4	US-09-041-886-18	Sequence 18, Appl
38	31.8	2.4	936	4	US-09-179-558-62	Sequence 62, Appl
39	31.8	2.4	24417	2	US-08-846-762-1	Sequence 1, Appl
40	31.6	2.4	1820	1	US-08-173-508-7	Sequence 7, Appl
41	31.6	2.4	1821	2	US-08-265-310-7	Sequence 7, Appl
42	31.6	2.4	1821	3	US-08-951-742-7	Sequence 7, Appl
43	31.2	2.3	2335	4	US-09-387-574-9	Sequence 9, Appl
44	31.2	2.3	2823	1	US-08-398-008A-1	Sequence 1, Appl
45	31.2	2.3	2823	2	US-08-893-333-1	Sequence 1, Appl

ALIGNMENTS

RESULT 1
US-09-120-817-1
; Sequence 1, Application US/09120817
; Patent No. 6172184
; GENERAL INFORMATION:
; APPLICANT: Collmer, Alan
; APPLICANT: Charkowski, Amy
; APPLICANT: Alfano, James R.
; TITLE OF INVENTION: HYPERSENSITIVE RESPONSE ELICITOR FROM
; TITLE OF INVENTION: PSEUDOMONAS SYRINGAE AND ITS USE
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Nixon, Hargrave, Devans & Doyle LLP
; CITY: Rochester
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 14603
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION NUMBER: US/09/120,817
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/055,107
; FILING DATE: 08-AUG-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Goldman, Michael L.
; REGISTRATION NUMBER: 30,727
; REFERENCE/DOCKET NUMBER: 19603/1741
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (716) 263-1304
; TELEFAX: (716) 263-1600
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1729 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-09-120-817-1

Query Match 11.5%; Score 155.2; DB 4; Length 1729;
Best Local Similarity 57.1%; Pred. No. 1e-38;
Matches 330; Conservative 0; Mismatches 233; Indels 15; Gaps 2;


```

Qy      874  ctgtttactactgaagacggtgtccagcgtctgaaaaacgtccaccatgtggcg 922
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      127  .atcttctgtctgaggtctgggccaegcctgaaaaatgtagtattcgtgtg 175

RESULT      5
US-08-232-463-14
; Sequence 14, Application US/082322463
; Patent No. 5670367
; GENERAL INFORMATION:
; APPLICANT: DORNER, F.
; APPLICANT: SCHEIFELINGER, F.
; APPLICANT: FALKNER, F. G.
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 1800 Diagonal Road, Suite 500
; CITY: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22313-0299
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk

```

SOFTWARE: PatentIn Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/232.463
 FILING DATE:
 CLASSIFICATION: 435
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US/07/935.313
 FILING DATE:
 APPLICATION NUMBER: EP 91 114 300.6
 FILING DATE: 26-AUG-1991
 ATTORNEY/AGENT INFORMATION:
 NAME: BENT, Stephen A.
 REGISTRATION NUMBER: 29,768
 REFERENCE/DOCKET NUMBER: 30472/114 IMMU
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (703)836-9300
 TELEFAX: (703)683-4109
 TELEX: 899149
 INFORMATION FOR SEQ ID NO: 14:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 7218 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 IMMEDIATE SOURCE:
 CLONE: PTZgpt-Fls
 US-08-232-463-14

Query Match	3.4%;	Score 45.8;	DB 1;	Length 7218;
Best Local Similarity	4.9%;	Pred. NO. 0.00054;		
Matches 11; Conservative 135; Mismatches 77; Indels 0; Gaps				
Qy	426	gatggacgcccgaagcgatcagcttggccaaacctggtacgggcacaacagtgcctcttc	485	
		: : : : : : : : : : : : : : :		
Db	1048	GGTCTGAGGAGCTTGGATYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY	1107	
Qy	486	cvgttactcttctcatctggcgggtttcccccttttaacgatctatcagggggaagggcccttc	545	
	: :			
Db	1108	YVYVV	1167	
Qy	546	cggcaactcccttcgccgaactactctctccgctcagtagtacccttctcaccccatcagcc	605	
	: :			
Db	1168	YVYVV	1227	
Qy	606	aacgtcccttacctcacccgcttgattctctctctctctctccacc	648	

RESULT 11
US-08-945-848-7
: Sequence 7, Application US/08945848
: Patent No. 5968772
: GENERAL INFORMATION:
: APPLICANT: MATSUSHIRO, Aizo
: TITLE OF INVENTION: PEARL PROTEIN(NACREIN) AND PROCESS FOR
: TITLE OF INVENTION: THE SAME
: NUMBER OF SEQUENCES: 9
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
: STREET: 233 South Wacker Drive/6300 Sears Tower
: CITY: Chicago

	Query Match	2.6%	Score 35.2;	DB 2;	Length 1341;
	Best Local Similarity	47.0%	Pred. No. 0.4;		
	Matches 109;	Conservative 0;	Mismatches 123;	Indels 0;	Gaps 0;
Qy	803	atgcaagacaaaccttcacgcggttcagaaatagcgatgagcgccagctctgaaa	862		
Db	716	ACGCANTGGAGACAATGGCAATAACGGCTACATGGGGACACCGTAAACATGGTGACA	775		
Qy	863	accgaaaaccgctgtttatactggaagacggtccagcctgaaaaacatcaccatggygcg	922		
Db	776	ACGSCAATAACAGCTACAATGGGGACACCGTAAACAATGGTGTCAACGGCAATAACGGCT	835		
Qy	923	acgacggggcggaatggtattcatctttacggtgatgcacaaatagacacatctgcacgcca	982		
Db	836	ACATGGGGACACCGGTAAACAATGGAGACACACGGCAATAACGGCTACANTGGGGACAACG	895		
Qy	983	ccaaactggttgagacgcgattaccgtttagcgcacacgcgcgggcaaaaa	1034		
Db	896	GTAAACAATGGTGACAACCGGCAATAACGGTGAACACGGCAATAACCGGTGAAAA	947		

RESULT 12
US-08-945-848-6
; Sequence 6, Application US/08945848
; Patent No. 5968772
; GENERAL INFORMATION:
; APPLICANT: MATSUISHI, Aizo
; TITLE OF INVENTION: PEARL PROTEIN(NACREIN) AND PROCESS FOR
; TITLE OF INVENTION: PEARL PROTEIN(NACREIN) AND PROCESS FOR
; TITLE OF INVENTION: THE SAME

NUMBER OF SEQUENCES: 9
CORRESPONDENCE ADDRESS:
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
STREET: 233 South Wacker Drive/6300 Sears Tower
CITY: Chicago
STATE: Illinois
COUNTRY: United States of America
ZIP: 60606
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/945,848
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Cawley, Jr, Thomas A.
REGISTRATION NUMBER: 40,944
REFERENCE/DOCKET NUMBER: 19036/34324
TELECOMMUNICATION INFORMATION:
TELEPHONE: (312) 474-6300
TELEFAX: (312) 474-0448
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 2363 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 22..1362
FEATURE:
NAME/KEY: sig_peptide
LOCATION: 22..72
FEATURE:
NAME/KEY: binding site
LOCATION: 73..120
FEATURE:
NAME/KEY: repeat_region
LOCATION: 745..990
FEATURE:
NAME/KEY: repeat_unit
LOCATION: 745..748
FEATURE:
NAME/KEY: active site
LOCATION: 1144..1287
US-08-945-848-6

Query Match 2.6%; Score 35.2; DB 2; Length 2363;
Best Local Similarity 47.0%; Pred. No. 0.57;
Matches 109; Conservative 0; Mismatches 123; Indels 0; Gaps 0;
QY 803 atggcaagagacaacattccacgcggttcagaaattagcgatgagcgccagctctgaaa 862
DB 737 ACGGCAATGGAGACATGGCAATAACGGCTACAAATGGGGACACGGTAACAATGGTGACA 796
QY 863 accagaaaccgctgtttatactggaagacggtgcccagcctgaaaaacgtcaccatggcg 922
DB 797 ACGGCAATAACAGCTACAATGGGGACACCGTTAACAATGGTGTCAACGGCAATAACGGCT 856
QY 923 acgacgggaggatgtattctattcttagcgtgacccaaatagcaaatctcagctca 982
DB 857 ACAATGGGGACACGGTAACAATGGAGACAACGGCAATAACGGCTACAATGGGGACAACG 916
QY 983 ccaacgtgggtgagcgcgattaccgtttaagccaaacgcgaggcaaaaa 1034
DB 917 GTAAATGGTGACACACGGCAATAACGGTGAACACGGCAATAACGGTGAAAA 968

RESULT 13
US-09-130-114-2
Sequence 2, Application US/09130114
Patent No. 5976807
GENERAL INFORMATION:
APPLICANT: Horlick, Robert A.
APPLICANT: Dammaj, Bassam B.
APPLICANT: Robbins, Alan K.
TITLE OF INVENTION: Eukaryotic Cells Stably Expressing Genes
FROM MULTIPLE TRANSFECTED EPISODES
FILE REFERENCE: 0867/1D903US1
CURRENT APPLICATION NUMBER: US/09/130,114
CURRENT FILING DATE: 1998-08-06
NUMBER OF SEQ ID NOS: 36
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 2
LENGTH: 1931
TYPE: DNA
ORGANISM: EBNA
US-09-130-114-2

Query Match 2.6%; Score 34.6; DB 2; Length 1931;
Best Local Similarity 47.8%; Pred. No. 0.78;
Matches 100; Conservative 0; Mismatches 109; Indels 0; Gaps 0;
QY 542 ctcccggaactcccttcggaactactctcccgctcagtagctctctcccccaccca 601
DB 868 ctcccgctccctcccgctccctccctcccgctccctccctccctccctccctcc 927
QY 602 cgccaacgtcccttaccacccgcttgattccctctctcccccacccaagcagcg 661
DB 928 cccgctccctccctccctccctccctccctccctccctccctccctccctcc 987
QY 662 gcagcgcgcggttaacgcatctctgacccctgttgtagcgggagatcgggccgaa 721
DB 988 ccagctctccgctcaccctccgcccagctccctccatcaactccgcccagctcc 1047
QY 722 attcgggtgcttcaccagcgccggcgct 750
DB 1048 tcacctccggggcccccatctctctgca 1076

RESULT 14
PCT-US95-05741-10/c
Sequence 10, Application PC/TUS9505741
GENERAL INFORMATION:
APPLICANT: Weintraub, Harold
APPLICANT: Lee, Jacqueline E.
APPLICANT: Tapscott, Stephen J.
APPLICANT: Hollenberg, Stanley M.
TITLE OF INVENTION: Neurogenic Differentiation (NeuroD) Gene
NUMBER OF SEQUENCES: 11
CORRESPONDENCE ADDRESS:
ADDRESSEE: Christensen O'Connor Johnson Kindness
STREET: 1420 Fifth Avenue, Suite 2800
CITY: Seattle
STATE: WA
COUNTRY: USA
ZIP: 98101-2347
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/05741
FILING DATE:
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Broderick, Thomas F.
REGISTRATION NUMBER: 31,332

```

RESULT 15
US-08-233-788A-58/c
: Sequence 58, Application US/08233788A
: Patent No. 5635617
: GENERAL INFORMATION:
: APPLICANT: Doran, James L.
: APPLICANT: Kay, William W.
: APPLICANT: Collinson, Karen S.
: APPLICANT: Clouthier, Sharon C.
: TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR DETECTION
: TITLE OF INVENTION: OF SALMONELLA
: NUMBER OF SEQUENCES: 61
: CORRESPONDENCE ADDRESS:
: ADDRESS: Seed and Berry
: STREET: 6300 Columbia Center, 701 Fifth Avenue
: CITY: Seattle
: STATE: Washington
: COUNTRY: U.S.A.
: ZIP: 98104-7092
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/233,788A
: FILING DATE: 26-APR-1994
: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:
: NAME: King, Joshua
: REGISTRATION NUMBER: 35,570
: REFERENCE/DOCKET NUMBER: 920043.403C2
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (206) 622-4900
: TELEFAX: (206) 682-6031
: TELEX: 3723836 SEEDANBERRY
: INFORMATION FOR SEQ ID NO: 58:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 456 base pairs

```


GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 21, 2001, 10:50:32 ; Search time 86.67 seconds
(without alignments)
9736.931 Million cell updates/sec

Title: US-09-596-958-1

Perfect score: 1344

Sequence: 1 atgtcaattcttaagcttaa.....acctgaagggtggatgaatga 1344

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 730101 seqs, 313950809 residues

Total number of hits satisfying chosen parameters: 1460202

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

N_Geneseq_0601:*

1: /SIDS1/gcgdata/geneseq/geneseq/NA1980.DAT:*

2: /SIDS1/gcgdata/geneseq/geneseq/NA1981.DAT:*

3: /SIDS1/gcgdata/geneseq/geneseq/NA1982.DAT:*

4: /SIDS1/gcgdata/geneseq/geneseq/NA1983.DAT:*

5: /SIDS1/gcgdata/geneseq/geneseq/NA1984.DAT:*

6: /SIDS1/gcgdata/geneseq/geneseq/NA1985.DAT:*

7: /SIDS1/gcgdata/geneseq/geneseq/NA1986.DAT:*

8: /SIDS1/gcgdata/geneseq/geneseq/NA1987.DAT:*

9: /SIDS1/gcgdata/geneseq/geneseq/NA1988.DAT:*

10: /SIDS1/gcgdata/geneseq/geneseq/NA1989.DAT:*

11: /SIDS1/gcgdata/geneseq/geneseq/NA1990.DAT:*

12: /SIDS1/gcgdata/geneseq/geneseq/NA1991.DAT:*

13: /SIDS1/gcgdata/geneseq/geneseq/NA1992.DAT:*

14: /SIDS1/gcgdata/geneseq/geneseq/NA1993.DAT:*

15: /SIDS1/gcgdata/geneseq/geneseq/NA1994.DAT:*

16: /SIDS1/gcgdata/geneseq/geneseq/NA1995.DAT:*

17: /SIDS1/gcgdata/geneseq/geneseq/NA1996.DAT:*

18: /SIDS1/gcgdata/geneseq/geneseq/NA1997.DAT:*

19: /SIDS1/gcgdata/geneseq/geneseq/NA1998.DAT:*

20: /SIDS1/gcgdata/geneseq/geneseq/NA1999.DAT:*

21: /SIDS1/gcgdata/geneseq/geneseq/NA2000.DAT:*

22: /SIDS1/gcgdata/geneseq/geneseq/NA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1344	100.0	1344	20 AAX09007	Hypersensitive res
2	1344	100.0	1344	21 AAD00669	Erwinia amylovora
3	1344	100.0	1344	21 AAD014939	DNA encoding a hyp
4	155.2	11.5	1729	20 AAX24347	Hypersensitive res
5	155.2	11.5	1729	21 AAD00673	Pseudomonas syring
6	155.2	11.5	1729	21 AAD014943	DNA encoding a hyp
7	94.8	7.1	591	19 AAV69879	Bacillus sp strain
8	94.8	7.1	591	19 AAV59478	Bacillus sp. pecti
9	74.2	5.5	666	20 AAX89484	Bacillus lichenifo
10	46.6	3.5	732	21 AAF12333	Aspergillus oryzae
11	42.8	3.2	1166	20 AAZ19440	M. tuberculosis an

12	42.8	3.2	1166	20	AAZ19228	M. tuberculosis re
13	38.2	2.8	1925	20	AAX90924	Epstein Barr Virus
14	37.6	2.8	4289	19	AAV62147	HSV-2 strain SB5 C
15	37.4	2.8	2371	14	AAQ50643	Human Ews gene clo
16	37.4	2.8	3309	14	AAQ50646	Human Ews gene (ge
17	37	2.8	117213	19	AAV62176	HSV-2 strain SB5 C
18	36.4	2.7	1320	22	AAF72217	Corynebacterium gl
19	36.2	2.7	400	19	AAV27417	Streptococcus pneu
20	36.2	2.7	3510	19	AAV52186	Streptococcus pneu
21	35.8	2.7	1370	21	AAV13346	Aspergillus oryzae
22	35.8	2.7	3453	21	AAV79709	Eucalyptus grandis
23	35.8	2.7	53526	19	AAV94101	Human PKD1 gene.
24	35.8	2.7	53577	17	AAV18551	Human polycystic k
25	35.8	2.7	53577	19	AAV94108	Human PKD1 locus b
26	35.8	2.7	114955	20	AAX53491	Human adenosine Al
27	35.4	2.6	335	21	AAV44637	Human secreted exp
28	35.2	2.6	2363	17	AAV58313	Oyster pearl prote
29	34.8	2.6	1335	21	AAV79118	Human cancer assoc
30	34.4	2.6	596	21	AAV14025	Aspergillus oryzae
31	34.4	2.6	6032	19	AAV30268	Plasmid pMT1802 en
32	34.4	2.6	6032	19	AAV12368	Trichoderma harzia
33	34.2	2.5	485	16	AAV05516	Human neurogenic d
34	34.2	2.5	1337	20	AAV17263	Human gene express
35	34.2	2.5	114955	20	AAX53491	Human adenosine Al
36	34	2.5	456	15	AAV87467	Agfa sequence. Sa
37	34	2.5	456	18	AAV74142	Salmonella enterit
38	34	2.5	456	21	AAV64617	Salmonella enterit
39	34	2.5	456	21	AAV64622	Agfa::PT3#1 DNA se
40	34	2.5	456	21	AAV64623	Agfa::PT3#2 DNA se
41	34	2.5	456	21	AAV64626	Agfa::PT3#5 DNA se
42	34	2.5	456	21	AAV64627	Agfa::PT3#6 DNA se
43	34	2.5	456	21	AAV64628	Agfa::PT3#7 DNA se
44	34	2.5	456	21	AAV64630	Agfa::PT3#9 DNA se
45	34	2.5	456	21	AAV64631	Agfa::PT3#10 DNA s

ALIGNMENTS

RESULT 1

AAX09007
ID AAX09007 standard; DNA; 1344 BP.

XX AC AAX09007;

XX DT 14-JUN-1999 (first entry)

XX DE Hypersensitive response eliciting protein coding sequence (hrpW).

XX KW Hypersensitive response elicitor protein; hypersensitive response;

XX KW hrpW; pathogen; infection; crop protection; disease resistance;

XX KW pest resistance; transgenic plant; colouration; maturation; ss.

XX OS Erwinia amylovora.

XX FH Key Location/Qualifiers

XX FT CDS 1..1344

XX FT /*tag= a

XX FT /product= Hypersensitive_response_elicitor_protein

XX PN WO9907208-A1.

XX PD 18-FEB-1999.

XX PF 27-JUL-1998; 98WO-US15547.

XX PR 06-AUG-1997; 97US-0055108.

XX PA (CORR) CORNELL RES FOUND INC.

XX PI Beer SV, Kim JF;

XX XX WPI; 1999-167126/14.

DR

DR P-PSDB; AAW96260.
XX New Erwinia amylovora hypersensitive response eliciting gene and
PT protein - useful for providing transgenic plants and seeds with
PT enhanced growth, and insect and disease resistance
XX
XX
PS Claim 1; Page 49; 54pp; English.
XX
CC The hypersensitive response eliciting protein (hrp) or polypeptide
CC is produced as part of an active defense by plants against
CC incompatible pathogen infections. The hypersensitive response is a
CC rapid localised necrosis. The hrp protein and gene when used in
CC nucleotide constructs are useful for providing disease resistance to
CC plants, insect control to plants, and enhancing plant growth
CC (enhancing fruit size and earlier colouration and maturation), by
CC direct application of the protein to plants, or by producing
CC transgenic plants or seeds using the hrp gene.
XX
SQ Sequence 1344 BP; 343 A; 372 C; 357 G; 272 T; 0 other;

Query Match 100.0%; Score 1344; DB 20; Length 1344;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1344; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 atgtcaattcttaccgtttaaacaacaataactctgtctctccggtgtctgttccagtcgag 60
DB 1 atgtcaattcttaccgtttaaacaacaataactctgtctctccggtgtctgttccagtcgag 60

QY 61 ggggacaacggcgttggtgataatgataatctgcttgggggcaaacacccatcgat 120
DB 61 ggggacaacggcgttggtgataatgataatctgcttgggggcaaacacccatcgat 120

QY 121 cggcaaacattgagcaaatggtcctaatattgctgggaactgttaaagtcactgtctatcg 180
DB 121 cggcaaacattgagcaaatggtcctaatattgctgggaactgttaaagtcactgtctatcg 180

QY 181 ccacaatacaggtaattcggcaacccggagcgggtggcaatgaccagactacaggagttggt 240
DB 181 ccacaatacaggtaattcggcaacccggagcgggtggcaatgaccagactacaggagttggt 240

QY 241 aacgctggcggcctgaacgagcaaaagacacagcaggaacacactccgactgtgacagt 300
DB 241 aacgctggcggcctgaacgagcaaaagacacagcaggaacacactccgactgtgacagt 300

QY 301 cagaacatgctgagtgagatgggcaacaacggcgtgagtcaggccatcacgcccgatggc 360
DB 301 cagaacatgctgagtgagatgggcaacaacggcgtgagtcaggccatcacgcccgatggc 360

QY 361 caggggcggcggcagatcggcgataatctcttactgaaagccaatgctgaagcttattgca 420
DB 361 caggggcggcggcagatcggcgataatctcttactgaaagccaatgctgaagcttattgca 420

QY 421 cgcgatgagcggccaaagcagatcagtttggccaacctggtacggggaacaacagtgc 480
DB 421 cgcgatgagcggccaaagcagatcagtttggccaacctggtacggggaacaacagtgc 480

QY 481 tcttcgggtacttcttctcattcgtcggttcccttttaacgatctatcagggggaagggc 540
DB 481 tcttcgggtacttcttctcattcgtcggttcccttttaacgatctatcagggggaagggc 540

QY 541 ccttcgggaactcccttcgggcaactactctcccgtaagtaacctctcaaccccatcc 600
DB 541 ccttcgggaactcccttcgggcaactactctcccgtaagtaacctctcaaccccatcc 600

QY 601 acgccaacatccctaccctcacgcttgatttcccttcttcccaacaaagcagcgagg 660
DB 601 acgccaacatccctaccctcacgcttgatttcccttcttcccaacaaagcagcgagg 660

QY 661 ggcagcacgcggtaaccgatcatctgacctgttggttagcgcggtcatcgggccgga 720
DB 661 ggcagcacgcggtaaccgatcatctgacctgttggttagcgcggtcatcgggccgga 720

QY 721 aattcgggtgaccttaccagcgccgctaaatcaagacggtgctgcatgacaccattacc 780
DB 721 aattcgggtgaccttaccagcgccgctaaatcaagacggtgctgcatgacaccattacc 780

QY 781 gtgaaagcgggtcaggtgttggatggcaaaagagacaaacottcaaccgccggttcagaaatta 840
DB 781 gtgaaagcgggtcaggtgttggatggcaaaagagacaaacottcaaccgccggttcagaaatta 840

QY 841 ggcgatggcggccagctctgaaacaccagaaacgcgtgtttatactggaagcgggtgccagc 900
DB 841 ggcgatggcggccagctctgaaacaccagaaacgcgtgtttatactggaagcgggtgccagc 900

QY 901 ctgaaaaacgctcaccatggcgacgacggggcggtatggtattcatctttacgggtgatgcc 960
DB 901 ctgaaaaacgctcaccatggcgacgacggggcggtatggtattcatctttacgggtgatgcc 960

QY 961 aaaaatagacaattctgcaactcaccacacgctgggtgagagcgcgtattaccgttaagccaaac 1020
DB 961 aaaaatagacaattctgcaactcaccacacgctgggtgagagcgcgtattaccgttaagccaaac 1020

QY 1021 agcgcgggcacaaaaatcccacgttgaaatcaactaacagttccttcgagcacgcctctgac 1080
DB 1021 agcgcgggcacaaaaatcccacgttgaaatcaactaacagttccttcgagcacgcctctgac 1080

QY 1081 aagatcctgcagctgaatgccgataactaacctgagcgttgacaacgtgaaggccaaagac 1140
DB 1081 aagatcctgcagctgaatgccgataactaacctgagcgttgacaacgtgaaggccaaagac 1140

QY 1141 ttgtgtaactttgtacgcaactaacggcgggtcacaacaggttaactgggtatctgaatctgagc 1200
DB 1141 ttgtgtaactttgtacgcaactaacggcgggtcacaacaggttaactgggtatctgaatctgagc 1200

QY 1201 catatcagcgcgagaagacggttaagtctctgcttcttctgtaaaagcgatagcaggggctaaac 1260
DB 1201 catatcagcgcgagaagacggttaagtctctgcttcttctgtaaaagcgatagcaggggctaaac 1260

QY 1261 gtcaatcaccagtgatatactcactgggtgatgttgaaacacactacaaaagtgcgagtcc 1320
DB 1261 gtcaatcaccagtgatatactcactgggtgatgttgaaacacactacaaaagtgcgagtcc 1320

QY 1321 gccaaactgaaggtgctgaatga 1344
DB 1321 gccaaactgaaggtgctgaatga 1344

RESULT 2
AAD00669
ID AAD00669 standard; DNA; 1344 BP.
XX
AC AAD00669;
XX
DT 08-SEP-2000 (first entry)
XX
DE Erwinia amylovora hypersensitive response elicitor encoding DNA #2.
XX
KW Hypersensitive response elicitor; environmental stress resistance;
plant; ds.
XX
XX Erwinia amylovora.
XX
FH Key Location/Qualifiers
CDS 1..1344
FT /*tag= a
FT /product= "Hypersensitive response elicitor"
XX
XX WO200028055-A2.
PN
XX 18-MAY-2000.
PD
XX 04-NOV-1999; 99WO-US26039.
PF
XX 05-NOV-1998; 98US-0107243.
PR
XX

PA (EDEN-) EDEN BIOSCIENCE CORP.

XX Wei Z, Schading RL;

XX WPI; 2000-376566/32.

XX P-PSDB; AAY1094.

XX Application of a hypersensitive response elicitor protein to plants to
PT Impart stress resistance -

XX Disclosures; Page 10; 84pp; English.

XX The patent discloses a method to impart stress resistance to plants by
CC applying a hypersensitive response elicitor in a non-infectious form to
CC a plant or seed. The present sequence is a DNA encoding hypersensitive
CC response elicitor protein from Erwinia amylovora. The protein is
CC heat stable, protease sensitive and suppressed by inhibitors of plant
CC metabolism. The present sequence is used to transform transgenic plant
CC or plant seeds to impart stress resistance.

XX Sequence 1344 BP; 343 A; 372 C; 357 G; 272 T; 0 other;

Query Match 100.0%; Score 1344; DB 21; Length 1344;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 1344; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 atgtcaatttcacgttaacaaacatacctcgtccctcgcgggtctgttccagtcggg 60
Db 1 atgtcaatttcacgttaacaaacatacctcgtccctcgcgggtctgttccagtcggg 60
Qy 61 ggggacacggcgttggtgcataatgcgaattctcgttggggcacaacaccatcgat 120
Db 61 ggggacacggcgttggtgcataatgcgaattctcgttggggcacaacaccatcgat 120
Qy 121 cggcaaacattagcaaaatggtcctaattattggcggaaactgttaaagtcactgctatcg 180
Db 121 cggcaaacattagcaaaatggtcctaattattggcggaaactgttaaagtcactgctatcg 180
Qy 181 ccacaatcaggtaatcgccgaacggagcgggtggcgaatgacagactacaggaagtgggt 240
Db 181 ccacaatcaggtaatcgccgaacggagcgggtggcgaatgacagactacaggaagtgggt 240
Qy 241 aacgctggcggcctgaacgagcaaaaggcacagcaggaacacactccgcagctgcagagt 300
Db 241 aacgctggcggcctgaacgagcaaaaggcacagcaggaacacactccgcagctgcagagt 300
Qy 301 cagaacatgctgagtgagatgggcaacaacggcgtggtgacagggccatcacgcccagtgcc 360
Db 301 cagaacatgctgagtgagatgggcaacaacggcgtggtgacagggccatcacgcccagtgcc 360
Qy 361 caggcggcggcagatcgcgataatccctttactgaagccatgctgaagcttatigca 420
Db 361 caggcggcggcagatcgcgataatccctttactgaagccatgctgaagcttatigca 420
Qy 421 cgcagatggagcggcgaagcagtcagtttgcccaacctggtacgggcaacaacagtgcc 480
Db 421 cgcagatggagcggcgaagcagtcagtttgcccaacctggtacgggcaacaacagtgcc 480
Qy 481 tcttcgggtacttcttcacatcgcggttcccttttaacagatctatcaggggggaagccc 540
Db 481 tcttcgggtacttcttcacatcgcggttcccttttaacagatctatcaggggggaagccc 540
Qy 541 ccttcgggcaactcccttcggcaactactctccctcagctactcttccacccatcc 600
Db 541 ccttcgggcaactcccttcggcaactactctccctcagctactcttccacccatcc 600
Qy 601 acgcaacgctcccttaccctaccgcttgatttcccttcttccacccacgaagcggcg 660
Db 601 acgcaacgctcccttaccctaccgcttgatttcccttcttccacccacgaagcggcg 660
Qy 661 ggcagacgcgcggtaacagatcatacctgacccctgttgtagcgcgggcacgcggga 720
Db 661 ggcagacgcgcggtaacagatcatacctgacccctgttgtagcgcgggcacgcggga 720

Db 661 ggcagacgcgcggtaacagatcatacctgacccctgttgtagcgcgggcacgcggga 720
Qy 721 aattcgggtgcttccacagcgcgcgtaataatcagacgggtgctgcatgacacattacc 780
Db 721 aattcgggtgcttccacagcgcgcgtaataatcagacgggtgctgcatgacacattacc 780
Qy 781 gtgaagcgggtcaggtgtttgatggcaaggacaaaccttcaccgcgggttcagaatta 840
Db 781 gtgaagcgggtcaggtgtttgatggcaaggacaaaccttcaccgcgggttcagaatta 840
Qy 841 ggcgatggcggcagctgtgaaacacagaaacccgtgtttatactggaagcgggtgcacgc 900
Db 841 ggcgatggcggcagctgtgaaacacagaaacccgtgtttatactggaagcgggtgcacgc 900
Qy 901 ctgaaaaacgtccacatggcgacgacggcggtggtatcttcttaccggtgagtcgc 960
Db 901 ctgaaaaacgtccacatggcgacgacggcggtggtatcttcttaccggtgagtcgc 960
Qy 961 aaaaatagacaatctgcagctcaccacacgtgggtgaggaacggtattaccgttaagccaaac 1020
Db 961 aaaaatagacaatctgcagctcaccacacgtgggtgaggaacggtattaccgttaagccaaac 1020
Qy 1021 agcggcggaacaaatcccaacgttgaaatacactaacagttccttcgagcagcctctgac 1080
Db 1021 agcggcggaacaaatcccaacgttgaaatacactaacagttccttcgagcagcctctgac 1080
Qy 1081 aagatcctgcagctgaaatgcgatacactaacctgagcgttgacaacggtgaaagccaaagac 1140
Db 1081 aagatcctgcagctgaaatgcgatacactaacctgagcgttgacaacggtgaaagccaaagac 1140
Qy 1141 ttggtactttgtacgcactaacggcggtcaacagggtaactgggactgaactctgagc 1200
Db 1141 ttggtactttgtacgcactaacggcggtcaacagggtaactgggactgaactctgagc 1200
Qy 1201 catatcgcgacgaagcggtaagtctctgttctttaaagcgtatagcaggggctgaac 1260
Db 1201 catatcgcgacgaagcggtaagtctctgttctttaaagcgtatagcaggggctgaac 1260
Qy 1261 gtcaatacagctgatactcactgggtgagttgaaacacactacaagtcggatgtcc 1320
Db 1261 gtcaatacagctgatactcactgggtgagttgaaacacactacaagtcggatgtcc 1320
Qy 1321 gccaaacctgaaggctgctgaatga 1344
Db 1321 gccaaacctgaaggctgctgaatga 1344

RESULT 3

AAAL4939

ID AAAL4939 standard; DNA; 1344 BP.

XX AC AAAL4939;

XX DT 08-AUG-2000 (first entry)

XX DE DNA encoding a hypersensitive response elicitor protein.

XX KW Hypersensitive response; insect control; disease resistance;

XX KW Hypersensitive response elicitor; plant growth; vegetable; crop;

XX OS Erwinia amylovora.

XX FH Key Location/Qualifiers

XX FT CDS 1..1344

XX FT /*tag= a

XX FT /product= "hypersensitive response elicitor protein"

XX PN WO200020452-A2.

XX PD 13-APR-2000.

XX XX 05-OCT-1999; 99WO-US23181.

XX Key Location/Qualifiers
 FT CDS 419..1693
 FT /*tag= a
 XX
 PN WO9907207-A1.
 XX
 PD 18-FEB-1999.
 XX
 XX 24-JUL-1998; 98WO-US15501.
 XX
 PF 06-AUG-1997; 97US-0055107.
 XX
 PR (CORR) CORNELL RES FOUND INC.
 XX
 PA Alfano JR, Charkowski A, Collmer A;
 XX
 XX WPI; 1999-167125/14.
 DR P-PSDB; AAW97851.
 XX
 XX New hypersensitive response eliciting (dsPE) gene and protein -
 PT useful for providing transgenic plants and seeds with enhanced
 PT growth, and insect and disease resistance
 XX
 PS Claim 1; Page 37-38; 56pp; English.
 XX
 CC This the DNA sequence of the dsPE gene of *Pseudomonas syringae*
 CC pv. tomato DC3000 that codes for a 42.9 kDa hypersensitive response
 CC elicitor (HRE) polypeptide (see AAW97851), termed hrpW. To identify
 CC any HRE-like genes in the *P. syringae* pv. tomato DS3000 DNA flanking
 CC hrpR, cosmid pCPP2357, which contains this region in vector pCPP47,
 CC was isolated. A series of subclones in pML123 were constructed and
 CC screened for 2 potential HR phenotypes: (1) the ability to promote
 CC tobacco HRE activity in *Pseudomonas fluorescens* cells carrying
 CC pCPP2274, a delhrpZ PHIRL derivative, and (ii) interference with
 CC the HRE activity of *P. fluorescens* cells carrying wild-type PHIRL.
 CC No subclones had the first phenotype, but one, pCPP2373, had the
 CC second. Transcriptional unit V of this subclone contained the
 CC 1275 bp open reading frame encoding HrpW. The HrpW protein or
 CC isolated DNA molecule can be used to impart disease resistance to
 CC plants, to enhance plant growth and/or to control insects on
 CC plants. This is achieved by applying the HRE protein in a
 CC non-infectious form to plants or plant seeds. Alternatively,
 CC transgenic plants or plant seeds transformed with DNA encoding the
 CC HRE can be provided.
 XX
 SQ Sequence 1729 BP; 464 A; 484 C; 486 G; 295 T; 0 other;

Query Match 11.5%; Score 155.2; DB 20; Length 1729;
 Best Local Similarity 57.1%; Pred. No. 9.6e-37;
 Matches 330; Conservative 0; Mismatches 233; Indels 15; Gaps 2;

QY 745 ggcgtcaatcagacggtgtgtatgacaccattaccgtgaaagcggtgtgtgtat 804
 DB 1079 gccgcaagatcaatgtgtgaaagaccattaccgtggtgtgtgtgtgtgtgt 1138
 QY 805 ggcagaggaacacccctccggtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 864
 DB 1139 ggcagaggaacacccctccggtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1198
 QY 865 cagaaacccgt 924
 DB 1199 cagaagccatgt 1258
 QY 925 gacggggcggt 972
 DB 1259 aacgaggt 1318
 QY 973 ctgcagctaccacacgt 1032
 DB 1319 gtgcgtccccagaacgt 1378

QY 1033 aatccacgttgaataactactaactagttctctgagcacgcctctgacaagactctgcag 1092
 DB 1379 gtcaataatctgaactcaagaacagcagtgccaaggttcagacagacaggtgtgtccag 1438
 QY 1093 ctgaatgccgataactaactcagcgtttgacaacgtgaaagccaaagacttctgtactttt 1152
 DB 1439 ctcaacgccaaactcacttgaatactgacaacttcaagccgacgatttctcgacgatg 1498
 QY 1153 gtacgcactaaccggtgtcaacag---ggttaactgggtatctgaatctgagccatatacgc 1209
 DB 1499 gttcgaccaaacggtgtggaagcagttttgatgacatgacatcgcagctgaacggtcgtgaa 1558
 QY 1210 gcagaagacggttaagttctctgttctgtaaaagcgatagcaggggttaaacgtcaatacc 1369
 DB 1559 gctaaccacggcaagttcgccctggtgaaagcgacagtgacgatctctgaagctggcaacg 1618
 QY 1270 agtgatatctcactcgtggt 1307
 DB 1619 ggcaacatgcctgacccgacgtcaaacacgcctctaga 1656

RESULT 5

AAD00673
 ID AAD00673 standard; DNA; 1729 BP.

XX
 AC AAD00673;

DT 08-SEP-2000 (first entry)

XX *Pseudomonas syringae* hypersensitive response elicitor encoding dsPE gene.

XX Hypersensitive response elicitor; environmental stress resistance;

XX plant; dsPE gene; ds.

XX *Pseudomonas syringae*.

XX Key Location/Qualifiers

FT CDS 419..1693

FT /*tag= a

FT /product= "Hypersensitive response elicitor"

XX WO200028055-A2.

XX 18-MAY-2000.

XX 04-NOV-1999; 99WO-US26039.

XX 05-NOV-1998; 98US-0107243.

XX (EDEN-) EDEN BIOSCIENCE CORP.

XX Wei Z, Schading RL;

XX WPI; 2000-376566/32.

XX P-PSDB; AAY71098.

XX Application of a hypersensitive response elicitor protein to plants to

XX impart stress resistance

XX Disclosure; Page 24-25; 84pp; English.

XX The patent discloses a method to impart stress resistance to plants by

XX applying a hypersensitive response elicitor in a non-infectious form to

XX a plant or seed. The present sequence is dsPE gene encoding

XX hypersensitive response elicitor protein from *Pseudomonas syringae*.

XX The present sequence is used to transform transgenic plant or plant

XX seeds to impart stress resistance.

XX Sequence 1729 BP; 464 A; 484 C; 486 G; 295 T; 0 other;

Query Match 11.5%; Score 155.2; DB 21; Length 1729;
 Best Local Similarity 57.1%; Pred. No. 9.6e-37;

Matches 330; Conservative 0; Mismatches 233; Indels 15; Gaps 2;

```
Qy 745 ggcgcataatcagacggtctgcatcacaccattaccgtgaaagcggtcaggtgtttat 804
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1079 gccgcagatcaatgtgtgaaagaccaccaaccaggctggcgagcgtctttgac 1138
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 805 ggcaaaagacaaacattcacccggttcagaaataggcgatggcgccagctctgaaac 864
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1139 ggcacgcgcgaacacctcacctgcgcacaaatctatgggtaacggagaccaggcgcaaat 1198
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 865 cagaaacgcgtgtttactctggaagacggtgccaagcctgaaacacgtcaccatggcgac 924
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1199 cagaagcccatgttcagctgagtgaaagcgctacgtttgaaatgtgaacctgggtgag 1258
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 925 gacggcggtggtgtatctatcttcaag-----gtgatgccaaatagacaaat 972
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1259 aacgaggtcgatggcatcacctgaaagcccaaaacgctcaggaagtcaccattgacaaac 1318
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 973 ctgcacgtcaccaacgctgggtgagacgctattaccgtttaagccaaacagcgcgccaaa 1032
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1319 gtgcacgtcccaagacgtcggtgaagacctgattacggtcacaagcgagggcgcgagc 1378
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1033 aaatcccacgttgaaatcacataacagttccttcgagacgcctctgacaagatcctgag 1092
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1379 gtcaactaatctgaacatcaagaacagcagtgccaaaggtgcagacgaaggtgtccag 1438
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1093 ctgaatgcgcataactaacctgacgcttgacaacgtgaagggccaaagactttgtacttt 1152
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1439 ctcaacgcacaacactcaacttgaataatcgaaactcaaggcgacgattcggcacgatg 1498
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1153 gtacgcataacgcgggtcaacag---ggtactgggtatctgaatctgagccatctcagc 1209
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1499 gtccgcaccaacgcgttcgcaagcagtttgatgacacgagcatcgagctgaacggcatcgaa 1558
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy - 1210 gcagaagcggtaagtctctgcttctgcttaaaagcgtatagcggggtctaaagctcaatacc 1269
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1559 gctaaccacggcaagttcgcctgtgtgaaagcgacagtgcagatctgaagctggcaacg 1618
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1270 agtgatatctcactgggtgatgttgaaacacactacaa 1307
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1619 ggcaacatcgccatgaccgacgtcaaacacgcctacga 1656
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

```
RESULT 6
ID AAA14943
ID AAA14943 standard; DNA; 1729 BP.
XX
AC AAA14943;
XX
DT 08-AUG-2000 (first entry)
XX
DE DNA encoding a hypersensitive response elicitor protein.
XX
KW Hypersensitive response; Insect control; disease resistance;
KW hypersensitive response elicitor; plant growth; vegetable; crop;
KW ornamental plant; ss.
XX
OS Pseudomonas syringae.
XX
FH Key Location/Qualifiers
FT CDS 419..1693
FT FT /*tag= a
FT FT /product= "hypersensitive response elicitor protein"
XX
PN W0200020452-A2.
XX
PD 13-APR-2000.
XX
PF 05-OCT-1999; 99WO-US23181.
XX
PR 05-OCT-1998; 98US-0103050.
XX
FA (EDEN-) EDEN BIOSCIENCE CORP.
```

```
XX Wei Z, Fan H, Niggemeyer JL;
PI WPT: 2000-303745/26.
DR P-FSDB: AAY84859.
XX
PT Hypersensitive response elicitor polypeptides useful for imparting
PT enhanced growth, disease resistance and insect resistance to plants,
PT especially vegetables and ornamental flowers -
XX
PS Disclosure: Page 25-26; 100pp; English.
XX
CC The present sequence encodes a hypersensitive response elicitor
CC polypeptide. The specification describes hypersensitive response
CC elicitor polypeptide fragments, which do not elicit a hypersensitive
CC response. Instead, the proteins impart disease resistance to plants,
CC enhance plant growth, and/or control insects. The polypeptide
CC fragments may be used to these properties to plants. The plants which
CC may be treated in this way include vegetables, crops and ornamental
CC plants such as alfalfa, rice, wheat, barley, rye, cotton, sunflower,
CC peanut, corn, potato, sweet potato, bean, pea, chicory, lettuce,
CC endive, cabbage, brussel sprout, beet, parsnip, turnip, cauliflower,
CC broccoli, radish, spinach, onion, garlic, eggplant, pepper, celery,
CC carrot, squash, pumpkin, zucchini, cucumber, apple, pear, melon,
CC citrus, strawberry, grape, raspberry, pineapple, soybean, tobacco,
CC tomato, sorghum or sugarcane, Arabidopsis thaliana, Saintpaulia,
CC petunia, pelargonium, poinsettia, chrysanthemum, carnation or zinnia.
XX
SQ Sequence 1729 BP; 464 A; 484 C; 486 G; 295 T; 0 Other;
```

```
Query Match 11.5%; Score 155.2; DB 21; Length 1729;
Best Local Similarity 57.1%; Pred. No. 9.6e-37;
Matches 330; Conservative 0; Mismatches 233; Indels 15; Gaps 2;
Qy 745 ggcgcataatcagacggtctgcatcacaccattaccgtgaaagcggtcaggtgtttat 804
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1079 gccgcagatcaatgtgtgaaagaccaccaaccaggctggcgagcgtctttgac 1138
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 805 ggcaaaagacaaacacctcacccggttcagaaataggcgatggcgccagctctgaaac 864
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1139 ggcacgcgcgaacacctcacctgcgcacaaatctatgggtaacggagaccaggcgcaaat 1198
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 865 cagaaacgcgtgtttactctggaagacggtgccaagcctgaaacacgtcaccatggcgac 924
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1199 cagaagcccatgttcagctgagtgaaagcgctacgtttgaaatgtgaacctgggtgag 1258
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 925 gacggcggtggtgtatctatcttcaag-----gtgatgccaaatagacaaat 972
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1259 aacgaggtcgatggcatcacctgaaagcccaaaacgctcaggaagtcaccattgacaaac 1318
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 973 ctgcacgtcaccaacgctgggtgagacgctattaccgtttaagccaaacagcgcgccaaa 1032
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1319 gtgcacgtcccaagacgtcggtgaagacctgattacggtcacaagcgagggcgcgagc 1378
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1033 aaatcccacgttgaaatcacataacagttccttcgagacgcctctgacaagatcctgag 1092
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1379 gtcaactaatctgaacatcaagaacagcagtgccaaaggtgcagacgaaggtgtccag 1438
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1093 ctgaatgcgcataactaacctgacgcttgacaacgtgaagggccaaagactttgtacttt 1152
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1439 ctcaacgcacaacactcaacttgaataatcgaaactcaaggcgacgattcggcacgatg 1498
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1153 gtacgcataacgcgggtcaacag---ggtactgggtatctgaatctgagccatctcagc 1209
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1499 gtccgcaccaacgcgttcgcaagcagtttgatgacacgagcatcgagctgaacggcatcgaa 1558
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1210 gcagaagcggtaagtctctgcttctgcttaaaagcgtatagcggggtctaaagctcaatacc 1269
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1559 gctaaccacggcaagttcgcctgtgtgaaagcgacagtgcagatctgaagctggcaacg 1618
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1270 agtgatatctcactgggtgatgttgaaacacactacaa 1307
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

Db 1619 ggcaacatcgccatgaccgacgtcaaacacgcctaaga 1656

RESULT 7

ID AAV69879 standard; DNA; 591 BP.

XX AAV69879;

XX 28-JAN-1999 (first entry)

XX *Bacillus* sp strain KSM-P15 pectic acid lyase encoding DNA.

KW *Bacillus* sp. strain KSM-P15; pectic acid lyase; protopectinase;
KW protopectin; polygalacturonic acid; detergent; cotton; surfactant;
KW cellulase; protease; bleaching agent; ds.

XX *Bacillus* sp.

XX Key Location/Qualifiers
FH 1..591
CDS /*tag= a

XX WO9845393-A2.

XX 15-OCT-1998.

XX 08-APR-1998; 98WO-JP01613.

XX 08-SEP-1997; 97JP-0242736.

XX 09-APR-1997; 97JP-0091142.

XX (KAOS) KAO CORP.

XX Hatada Y, Ito S, Kasai M, Kobayashi T, Koike K;

XX Shikata S, Suzumatsu A, Tsumadori M, Wada Y;

XX WPI; 1998-568339/48.

XX P-PSDB; AAW83014.

XX Detergent composition containing protopectinase active at alkaline
PT pH - on protopectin and polygalacturonic acid, provides better
PT removal of muddy soil

XX Example; Page 71-72; 80pp; English.

XX The present invention describes a detergent composition which contains
CC a protopectinase having an optimum pH 7 or higher against protopectin
CC and polygalacturonic acid substrates. Inclusion of protopectinase
CC gives a composition that provides better removal of muddy soil,
CC particularly from socks. The present sequence encodes pectic acid lyase
CC from *Bacillus* sp. strain KSM-P15, which is used in an example from
CC the present invention. Pectic acid lyase exhibits protopectinase
CC activity.

XX Sequence 591 BP; 160 A; 134 C; 179 G; 118 T; 0 other;

Query Match 7.1%; Score 94.8; DB 19; Length 591;

Best Local Similarity 52.8%; Pred. No. 1.le-18;

Matches 263; Conservative 0; Mismatches 217; Indels 18; Gaps 2;

QY 754 cagacggtgctgcatacacaccattaccgtgaagcgggtcaggtgttgatgcaaaagga 813

Db 4 ccgaggtcgtctatgaacagattcgtgctcggtcagacgcttgcggaaggg 63

QY 814 caaaccttaccgcgggttcaga---attagcgatggcgccagctgtaaacaccagaaa 870

Db 64 cagacctatgtgctaataccgaatacatattggggacgagatcgagcgaggaatcagaag 123

QY 871 ccgctgtttatcagaagacggtgccagcctgaaaaagtcaccatggggcagcaggg 930

Db 124 ccgattcttctgtaggggtggggcgaagcctgaaaaatgattgattggcgtctcgc 183

QY 931 gcggatggtattcatctttacggtgatgccaaaatagacaatctgcagctcaccacggtg 990
Db 184 gctgacgggggtgctacgtcctacgggattgtacgattacaatgtcatctggggagatgtt 243
QY 991 ggtgagacgcgattaccgtttaagccaaacagcgcgggcaaaaaatcccacgttgaatc 1050
Db 244 ggtgaggtcgcgtgacgttaaatcgtccggaacg-----gtgaacatc 288
QY 1051 actaacagttccttcgagcgcgcctctgacaagatcctgcagctggaatgcgcgatactaac 1110
Db 289 tcggcggggcagcctcaaacaggcgtatgacaagggttcccaaatcaatgacgaggggacg 348
QY 1111 ctgagcgttgacaacgagcgaagccaaagactttggtacttttgcacgactaacgcggcgt 1170
Db 349 atcaacattcgtacttcaggcgcgatgacatcgggaagcgtgcttcggcgaacgaggc 408
QY 1171 caacgggttaactgggatcgtgaatctgagccatatcagcgcagaagaagcgttaagttctcg 1230
Db 409 accacctacaagtggtgatgaacgtggaaaactgcaacatttccagagtgaaggatgcg 468
QY 1231 ttggttaaaagcgatagc 1248
Db 469 atcctgagaacggacagc 486

RESULT 8

AAV59478

ID AAV59478 standard; DNA; 591 BP.

XX AAV59478;

XX 07-JAN-1999 (first entry)

XX *Bacillus* sp. pectic acid lyase encoding DNA.

XX *Bacillus* sp. KSM-P15; pectic acid lyase; pectinic acid lyase; pectin;
XX detergent; food-processing; fibre-processing agent; cell wall; ds.

XX *Bacillus* sp.

XX Key Location/Qualifiers
FH 1..591
CDS /*tag= a
FT /note= "no stop codon given"

XX EP870834-A1.

XX 14-OCT-1998.

XX 09-APR-1998; 98EP-0106586.

XX 08-SEP-1997; 97JP-0242735.

XX 09-APR-1997; 97JP-0091142.

XX (KAOS) KAO CORP.

XX Hatada Y, Ito S, Kobayashi T, Koike K, Suzumatsu A;

XX Yoshimatsu T;

XX WPI; 1998-523159/45.

XX P-PSDB; AAW77412.

XX New *Bacillus* pectic acid lyase - useful as a detergent component, a
PT food-processing agent and a fibre-processing agent

XX Claim 1; Page 16-17; 29pp; English.

XX The present sequence encodes a pectic acid lyase isolated from
CC microorganism *Bacillus* sp. KSM-P15. The pectic acid lyase has high
CC pectic acid lyase activity which degrades pectin in plant cell walls
CC and fibre in vegetables, and so is useful as a component of detergents,
CC a food-processing agent, or a fibre-processing agent. The pectic acid

	Query Match	3.2%	Score 42.8;	DB 20;	Length 1166;
	Best Local Similarity	28.6%;	Pred. NO. 0.0086;		
	Matches 68;	Conservative	55;	Mismatches 115;	Indels 0; Gaps
Qy	448	t t g g c c a a c c t t g g t a c g g g c a a c a c a g t g c t c t t c c g g t a c t t c t c a t t c a t t c g g c g t	507		
		: : : : :		: : : : :	
Db	563	t y t n n c c a m e s c t c t m t c m a a c t c k c c g g y c k e n m y c t e k c c a y n m a a c k k y	622		
Qy	508	t c c c c t t t t a a g a t c a t c a g g g g g a a g g c c c t t c o g g a a c t c c c c t t c o g g a a c	567		
		: : : : : : : : : : : : : : : : :		: : : : :	
Db	623	c y w e n y c m y c k e k a g w y k n m c t c w a c t c t m y n t t t c t c t e n k c c m k a c k n t t c t	682		
Qy	568	t a c t c c c c t a g t a c t t c t c a c c c c a t c a c a c g c c c c t a c t c a c c g c t	627		
		: : : : : : : : : : : : : : :		: : : : :	
Db	683	c w e s p c c c c a c a k a n y n c y a w m t m c m c t e k a e c c c y v n n y c c m m n c w m t c w t c l w	742		
Qy	628	g a t t c c c t t c t t c t c c a c c a a g c a g c o g g g g g c a g c a c g g t a c c a c g a t c a t c	685		
		: : : : : : : : : : : : : :		: : : : :	
Db	743	n a k a n c n t t c t c t c m m y m l m a c k c w c n n t c n c k s g a c y c t c t a c t k m k c c k m	800		

RESULT 12	
AAZ19228	
ID	AAZ19228 standard; cDNA; 1166 BP.
XX	
AC	AAZ19228;
XX	
DT	05-NOV-1999 (first entry)
XX	
DE	- M. tuberculosis recombinant antigen cDNA encoding 3' Erdsen-8.
XX	
KW	Antigen; diagnosis; detection; infection; antibody; immunisation;
KW	vaccine; immunity; ss.
XX	
OS	- Mycobacterium tuberculosis:
XX	
PN	W09942118-A2.
XX	
PD	26-AUG-1999.
XX	
PF	17-FEB-1999; 99WO-US03265.
XX	
PR	05-MAY-1998; 98US-0072596.
XX	
XX	18-FEB-1998; 98US-0024753.
XX	
PA	(CORI-) CORIXA CORP.
XX	
PI	Campos-Neto A, Dillon DC, Hendrickson RC, Houghton R;
PI	Lodes MJ, Reed SG, Skeiky YAW, Twardzik DR, Vedvick TS;
XX	
XX	WPI; 1999-527416/44.
XX	
XX	New polypeptide comprising antigenic portions of M. tuberculosis
PT	
XX	Claim 4; Page 304; 323pp; English.
PS	
XX	
CC	This invention describes novel recombinant antigens and their encoding
CC	nucleic acids derived from Mycobacterium tuberculosis. The novel
CC	polypeptides are useful for detecting M. tuberculosis infection in a
CC	biological sample by detecting antibodies which bind with the
CC	polypeptides, and are useful as vaccines for immunizing against
CC	M. tuberculosis infection. The new detection methods are needed as
CC	current vaccination strategies do not provide 100% immunity.
XX	
XX	Sequence 1166 BP; 117 A; 371 C; 121 G; 207 T; 350 other;
SQ	

Db	563	tytnncamcmstcttmtmaactctkcccggyckcnmyctctctckccaynmaacckkty	622
Qy	508	tcctcttttaacgatctatcagggggaagccctcttcggcgaactcccttcggaac	567
Db	623	cywcnwycmyckckcagwyknmctcwaactctmyntttctctcnkccmkacknttct	682
Qy	568	tactctccgtcagtaactcttcacccccatccacgcgaacagctccctctacatcacccggtt	627
Db	683	cwscscscscacacaymcyawcmntcmctckscscscycnnycmmmcwcmctcwtw	742
Qy	628	gattctccctttcttccaccacaagcagccgggggagcagcagccgggaacacgatactc	685
Db	743	nakcancntttctctctctmymtmackcwnntcnccksgaccytctcactkmkckcm	800
RESULT	13		
AAAX90924/c	ID	AAAX90924 standard; DNA; 1925 BP.	
XX	XX	AAAX90924;	
AC	XX		
XX	17-JAN-2000	(first entry)	
DT	XX		
XX	XX	Epstein Barr Virus Nuclear Antigen 1 (EBNA 1) DNA.	
DE	XX		
XX	KW	Epstein Barr Virus Nuclear Antigen 1 DNA; plasmid pCMVBENA; EBNA 1;	
KW	KW	episome; transfection; origin of replication; EBV orip; receptor;	
KW	KW	eucaryotic host cell; recombinant cell line; ion channel; gene therapy;	
KW	KW	multiple gene expression; transporter protein; transcription factor;	
KW	KW	adhesion molecule; antisense therapy; gene amplification;	
KW	KW	cell immortalisation; ds.	
XX	XX		
OS	XX	Epstein-barr virus.	
XX	XX		
FH	Key	Location/Qualifiers	
FT	CDS	1..1925	
FT	FT	/*tag= a	
FT	FT	/product= "EBNA 1"	
FT	FT	/transl_except= (pos:799..800, aa:Gly)	
FT	FT	/note= "The sequence is described throughout the	
FT	FT	specification as being 1926 nucleotides long, but a	
FT	FT	sequence of only 1925 bp has been given in figure 2"	
XX	XX		
PN	W09947647-A1.		
XX	XX		
PD	23-SEP-1999.		
XX	XX		
PF	12-FEB-1999;	99WO-US03307.	
XX	XX		
PR	18-MAR-1998;	98US-0040961.	
PR	06-AUG-1998;	98US-0130114.	
XX	XX		
PA	(PHAR-) PHARMACOPEIA INC.		
XX	XX		
PI	Damaj BB, Horlick RA, Robbins AK;		
XX	XX		
DR	WPI; 1999-610610/52.		
DR	P-PSDB; AAY28843.		
XX	XX		
PT	New method for expressing genes from recombinant eukaryotic cells,		
PT	useful for gene therapy		
XX	XX		
XX	Claim 24; Fig 2; 86pp; English.		
XX	XX		
CC	CC	The present sequence is a DNA encoding Epstein Barr Virus Nuclear	
CC	CC	Antigen 1 (EBNA 1), which is obtained from commercially available	
CC	CC	plasmid pCMVBENA. EBNA 1 protein is used to stably maintain episomes	
CC	CC	containing EBV origin of replication (orip) and a gene encoding	
CC	CC	protein or RNA of interest. Eucaryotic host cells expressing EBNA 1	
CC	CC	protein are transfected with these episomes to produce recombinant	
CC	CC	cell lines expressing multiple genes of interest. This provides a	
CC	CC	rapid and reliable method of stably expressing multiple genes in	
CC	CC	transfected cells. The episomes are useful in the transfection of genes	

PI Plougastel B, Thomas G, Zucman

Id
plong

```
XX WPI; 1993-386580/48.
DR P-PSDB; AAR44555.
XX
PT New nucleic acid of EWS gene and its hybrid(s) - contg. gene
sequence involved in chromosomal trans-location, also derived
PT mRNA, probes, fusion proteins etc., for diagnosis and treatment
PT of Ewing sarcoma and melanoma
XX
PS Disclosure; Fig 6; 123pp; French.
XX
CC The probes 22RR3 and 22RR12 were used to screen a human foetal
CC brain cDNA library (Stratagene cat.# 936206). The clone BFIAC5 was
CC identified and sequenced. It represents the entire coding region
CC and 3'-UTR of the Ews gene.
XX
SQ Sequence 2371 BP; 639 A; 587 C; 659 G; 486 T; 0 Other;

Query Match      2.8%; Score 37.4; DB 14; Length 2371;
Best Local Similarity 47.0%; Pred. No. 0.54;
Matches 116; Conservative 0; Mismatches 131; Indels 0; Gaps 0;

QY 534 gaagcccttcggcaactcccttcggcaactactctccgctcagtcacattctcacc 593
   || || || || || || || || || || || || || || || || || || || ||
DB 1969 GACGGTGTCTCGCTTTATCCATTTTCCAGGTCTCCACGTCCTCTCTCTCTCCCA 1910

QY 594 cccatccacgccaacgtccctacctcaccgcttgatttccctctcttccaccacaaagc 653
   || || || || || || || || || || || || || || || || || || || ||
DB 1909 TCTGTTCATCAAGGTCTCAGGGGGCCCCCAGGGCCACCTCGTCTCTCCACCAAGC 1850

QY 654 agccgggggagcagccggttaaccgatactcctgacctgttgtagcgcgggcaccgg 713
   || || || || || || || || || || || || || || || || || || || ||
DB 1849 CACCTCGGTCCATGCCCGGCCACCAAGGACGACCTCTGTCTCCACCCAGCGCCACCTC 1790

QY 714 ggcggaaattcgggtggccttcaccagccgcgctaatacagaggggtgctgatgacac 773
   || || || || || || || || || || || || || || || || || || || ||
DB 1789 TGAACATTCCACCGGGACCAACCATCATGAGGCCACCTCTCTCTCCCGCATGCCAC 1730

QY 774 cattacc 780
   || || ||
DB 1729 CAGGGCC 1723
```

Search completed: October 21, 2001, 11:33:28
Job time: 2576 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 21, 2001, 10:45:17 ; Search time 992.24 Seconds
(without alignments)
12803.994 Million cell updates/sec

Title: US-09-596-958-1
Perfect score: 1344
Sequence: 1 atgtcaattcttaagcttaa.....acctgaagggtgctgaatga 1344

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 10228115 seqs, 4726426750 residues 20456230
Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

EST:*

1:	gb_est1:*	44:	em_esthum10:*
2:	gb_est2:*	45:	em_esthum11:*
3:	gb_est3:*	46:	em_esthum12:*
4:	gb_est4:*	47:	em_esthum13:*
5:	gb_est5:*	48:	em_esthum14:*
6:	gb_est6:*	49:	em_esthum15:*
7:	gb_est7:*	50:	em_esthum16:*
8:	gb_est8:*	51:	em_esthum17:*
9:	gb_est9:*	52:	em_esthum18:*
10:	gb_est10:*	53:	em_esthum19:*
11:	gb_est11:*	54:	em_esthum20:*
12:	gb_est12:*	55:	em_esthum21:*
13:	gb_est13:*	56:	em_esthum22:*
14:	gb_est14:*	57:	em_esthum23:*
15:	gb_est15:*	58:	em_esthum24:*
16:	gb_est16:*	59:	em_esthum25:*
17:	gb_est17:*	60:	em_esthum26:*
18:	gb_est18:*	61:	em_esthum27:*
19:	gb_est19:*	62:	em_esthum28:*
20:	gb_est20:*	63:	em_estin1:*
21:	gb_est21:*	64:	em_estin2:*
22:	gb_est22:*	65:	em_estin3:*
23:	gb_est23:*	66:	em_estin4:*
24:	gb_est24:*	67:	em_estin5:*
25:	gb_est25:*	68:	em_estom1:*
26:	gb_est26:*	69:	em_estom2:*
27:	gb_est27:*	70:	em_estov1:*
28:	gb_est28:*	71:	em_estov2:*
29:	gb_est29:*	72:	em_estpl1:*
30:	gb_est30:*	73:	em_estpl2:*
31:	gb_est31:*	74:	em_estpl3:*
32:	gb_est32:*	75:	em_estpl4:*
33:	gb_est33:*	76:	em_estpl5:*
34:	gb_est34:*	77:	em_estpl6:*
35:	gb_est35:*	78:	em_estpl7:*
36:	gb_est36:*	79:	em_estpl8:*
37:	gb_est37:*	80:	em_estpl9:*
38:	gb_est38:*	81:	em_estpl10:*
39:	gb_est39:*	82:	em_estrol1:*
40:	gb_est40:*	83:	em_estrol2:*
41:	gb_est41:*	84:	em_estrol3:*
42:	gb_est42:*	85:	em_estrol4:*
43:	gb_est43:*	86:	em_estrol5:*
		87:	em_estrol6:*
		88:	em_estrol7:*
		89:	em_estrol8:*
		90:	em_estrol9:*
		91:	em_estrol10:*
		92:	em_estrol11:*
		93:	em_estrol12:*
		94:	em_estrol13:*
		95:	em_estrol14:*
		96:	em_estrol15:*
		97:	em_estrol16:*
		98:	em_estrol17:*
		99:	em_estrol18:*
		100:	em_estrol19:*
		101:	em_estrol20:*
		102:	gb_est25:*
		103:	gb_est26:*
		104:	gb_est27:*
		105:	gb_est28:*
		106:	gb_est29:*
		107:	gb_est30:*
		108:	gb_est31:*
		109:	gb_est32:*
		110:	gb_est41:*
		111:	gb_est42:*
		112:	gb_est43:*
		113:	gb_est44:*
		114:	gb_est45:*
		115:	gb_est46:*
		116:	gb_est47:*

117: gb_est48:*
118: gb_est49:*
119: gb_est50:*
120: gb_est51:*
121: gb_est52:*
122: gb_est53:*
123: gb_est54:*
124: gb_est55:*
125: gb_est56:*
126: gb_est57:*
127: gb_est58:*
128: gb_est59:*
129: gb_est60:*
130: gb_est61:*
131: gb_est62:*
132: gb_est63:*
133: gb_est64:*
134: gb_est65:*
135: gb_est66:*
136: gb_est67:*
137: gb_est68:*
138: gb_est69:*
139: gb_est70:*
140: gb_est71:*
141: gb_est72:*
142: gb_est73:*
143: gb_est74:*
144: gb_est75:*
145: gb_est76:*
146: gb_est77:*
147: gb_est78:*
148: gb_est79:*
149: gb_est80:*
150: gb_est81:*
151: gb_est82:*
152: gb_est83:*
153: gb_est84:*
154: gb_est85:*
155: gb_est86:*
156: gb_est87:*
157: gb_est88:*
158: gb_est89:*
159: gb_est90:*
160: gb_est91:*
161: gb_est92:*
162: gb_est93:*
163: gb_est94:*
164: gb_est95:*
165: gb_est96:*
166: gb_est97:*
167: gb_est98:*
168: gb_est99:*
169: gb_est100:*
170: gb_est101:*
171: gb_est102:*
172: gb_est103:*
173: gb_est104:*
174: gb_est105:*
175: gb_est106:*
176: gb_est107:*
177: gb_est108:*
178: gb_est109:*
179: gb_est110:*
180: gb_est111:*
181: gb_est112:*
182: gb_est113:*
183: gb_est114:*
184: gb_est115:*
185: gb_est116:*
186: gb_est117:*
187: gb_est118:*
188: gb_est119:*
189: gb_est120:*

190: gb_est121:*
191: gb_est122:*
192: gb_est123:*
193: gb_est124:*
194: gb_est125:*
195: gb_est126:*
196: gb_est127:*
197: gb_est128:*
198: gb_est129:*
199: gb_est130:*
200: gb_est131:*
201: gb_est132:*
202: gb_est133:*
203: gb_est134:*
204: gb_est135:*
205: gb_est136:*
206: gb_est137:*
207: gb_est138:*
208: gb_est139:*
209: gb_est140:*
210: gb_est141:*
211: gb_est142:*
212: gb_est143:*
213: gb_est144:*
214: gb_est145:*
215: gb_est146:*
216: gb_est147:*
217: gb_est148:*
218: gb_est149:*
219: gb_est150:*
220: gb_est151:*
221: gb_est152:*
222: gb_est153:*
223: gb_est154:*
224: gb_est155:*
225: gb_est156:*
226: gb_est157:*
227: gb_est158:*
228: gb_est159:*
229: gb_est160:*
230: gb_est161:*
231: gb_est162:*
232: gb_est163:*
233: gb_est164:*
234: gb_est165:*
235: gb_est166:*
236: gb_est167:*
237: gb_est168:*
238: gb_est169:*
239: gb_est170:*
240: gb_est171:*
241: gb_est172:*
242: gb_est173:*
243: gb_est174:*
244: gb_est175:*
245: gb_est176:*
246: gb_est177:*
247: gb_est178:*
248: gb_est179:*
249: gb_est180:*
250: gb_est181:*
251: gb_est182:*
252: gb_est183:*
253: gb_est184:*
254: gb_est185:*
255: gb_est186:*
256: gb_est187:*
257: gb_est188:*
258: gb_est189:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	63.8	4.7	469	152	CG310823	ro43d10.y
2	47	3.5	910	219	CNS0060N	AL065629 Drosophila
3	43.8	3.3	959	219	CNS008M2	AL052079 Drosophila
4	43.4	3.2	844	219	CNS0052P	AL056652 Drosophila
5	42.8	3.2	782	80	BF280877	GA_Eb004
6	42.8	3.2	782	153	BG447322	GA_Eb004
7	41.8	3.1	1260	170	BF866129	963066G03
8	41.6	3.1	1360	105	AL514267	AL514267
9	41.6	3.1	870	165	BE258873	601116824
10	41.6	3.1	975	220	CNS02KBD	AL201298 Tetraodon
11	41.2	3.1	391	152	CG310856	BG310856 ro44a03.y
12	41.2	3.1	495	17	A1201143	AI201143 qf64g10.x
13	41.2	3.1	534	112	AW173078	AW173078 xj82h09.x
14	41.2	3.1	563	19	AI336864	AI336864 qx86h08.x
15	41	3.1	467	106	AL566523	AL566523
16	40.8	3.0	669	228	AQ447577	AQ447577 mgxb0008E
17	40.8	3.0	977	219	CNS00JX7	AL076850 Drosophila
18	40.4	3.0	812	220	CNS02JGO	AL200193 Tetraodon
19	40.2	3.0	522	111	AW090025	AW090025 xc89d09.x
20	40.2	3.0	551	11	AA771081	AA771081 vt16c12.r
21	40.2	3.0	990	221	CNS03CU0	AL238257 Tetraodon
22	40.2	3.0	1101	219	CNS00EDF	AL068951 Drosophila
23	40	3.0	390	14	AA960471	AA960471 vw63a05.s
24	39.8	3.0	667	168	BF700488	BF700488 602128842
25	39.8	3.0	779	227	AQ331220	AQ331220 nbxb0049E
26	39.8	3.0	860	153	BG394463	BG394463 602456990
27	39.6	2.9	539	251	A2909203	A2909203 RPCI-24-1
28	39.6	2.9	793	255	B18549	B18549 t5A19-S66.T
29	39.6	2.9	995	219	CNS018MR	AL105533 Drosophila
30	39.6	2.9	1072	221	CNS03JEN5	AL240602 Tetraodon
31	39.4	2.9	878	137	BE572645	BE572645 601329421
32	39.2	2.9	968	147	BF341280	BF341280 602013214
33	39.2	2.9	1095	220	CNS02762	AL184292 Tetraodon
34	39	2.9	560	120	AW732191	AW732191 ba06f10.x
35	39	2.9	619	222	FR0007731	Z91541 F.rubripes
36	39	2.9	667	108	AU144606	AU144606 AU144606
37	39	2.9	856	220	CNS02HBF	AL197412 Tetraodon
38	39	2.9	997	168	BF683247	BF683247 602139267
39	39	2.9	1101	219	CNS002B1	AL097447 Drosophila
40	38.8	2.9	398	152	BG356477	BG356477 EMI_23_A0
41	38.8	2.9	455	137	BE601363	BE601363 P11_92_A1
42	38.8	2.9	568	154	BG463697	BG463697 EMI_50_H0
43	38.8	2.9	573	119	AW679673	AW679673 WSI_30_E0
44	38.8	2.9	597	152	BG323139	BG323139 EMI_15_D0
45	38.8	2.9	626	152	BG355747	BG355747 EMI_19_A0

ALIGNMENTS

RESULT	1
BG310823	ro43d10.y1 Heterodera glycines J2 pAMP1 v2.1 Chiapelli McCarter
LOCUS	Heterodera glycines cDNA 5' similar to TR:Q9X622 Q9X622 PEPTATE
DEFINITION	LYASE PRECURSOR ; mRNA sequence.
ACCESSION	BG310823
VERSION	BG310823.1 GI:13112624
KEYWORDS	EST.
SOURCE	Heterodera glycines.
ORGANISM	Heterodera glycines
REFERENCE	1 (bases 1 to 469)
AUTHORS	McCarter, J., Clifton, S., Chiapelli, B., Pape, D., Martin, J., Wylie, T., Dante, M., Marra, M., Hillier, L., Kucaba, T., Theising, B., Bowers, Y., Gibbons, M., Ritter, E., Bennett, J., Franklin, C., Tsagareishvili, R.

TITLE
JOURNAL
COMMENT

, Ronko, I., Kennedy, S., Maguire, L., Beck, C., Underwood, K., Steptoe, M., Allen, M., Person, B., Swaller, T., Harvey, N., Schurk, R., Kohn, S., Shin, T., Jackson, Y., Cardenas, M., McCann, R., Waterston, R. and Wilson, R.
The Washington Univ. Nematode EST Project, 1999
Unpublished (1999)
Contact: McCarter JP
The Washington Univ. Nematode EST Project, 1999
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@wustl.edu
The library was constructed by Brandi Chiapelli and Dr. James McCarter (bchiapelli@wustl.edu & jmcarter@wustl.edu) at Washington University, St. Louis. DNA sequencing by: Washington University Genome Sequencing Center St. Louis.
Seq primer: custom 5' primer (PAMP1_5'_{#7})
High quality sequence stop: 455.

FEATURES
Source

Location/Qualifiers
1..469
/organism="Heterodera glycines"
/db_xref="taxon:51029"
/clone_lib="Heterodera glycines J2 pAMP1 v2.1 Chiapelli McCarter"
/dev_stage="enriched for 2nd stage juveniles"
/lab_host="DH10B"
/note="vector: pAMP1; The library was constructed by Brandi Chiapelli and Dr. James McCarter at Washington University, St. Louis. The cDNA was made by using Dynabead oligo-dT priming (Dyna). PCR based library using a modified protocol from the SMART PCR cDNA Synthesis Kit from Clontech. Directionally cloned into the UDG sites of pAMP1. Nematodes are the OP50 strain developed by Dr. Charlie Opperman of North Carolina State University (see Genetics, 146:1311-8, 1997). Frozen J2 nematodes were provided by Dr. Rick Davis also of NCSU. Special thanks to Dr. Yulia Korshunova from the Lovett Lab at Washington University in St. Louis for helping with methods."

BASE COUNT
ORIGIN

144 a 96 c 124 g 104 t 1 others
Query Match 4.7%; Score 63.8; DB 152; Length 469;
Best Local Similarity 53.5%; Pred No. 2.6e-07;
Matches 154; Conservative 0; Mismatches 133; Indels 1; Gaps 1;
QY 815 aaaccttcacgcggttcagaattaggcgatggcgccagtcgtgaaaccagaaaccgc 874
Db 182 AACGGTTGACCGCTCTTCGGCATTTGGCACCGGTGATCAGTCGGAGGCGCAAAAGCAC 241
QY 875 tgtttatactggaagacgggtggcagcctgaaaaacgctcaccatggcgagcagcgggcg 934
Db 242 TGATCGGTGCGCAAAAGAAATGCAGTGTATCAAAAATTTGATCATTTGCCCAAAACGGCGG 301
QY 935 atggtattcatcttaccgtgtgatgccaaaatagacaatctg-cacgtccacaacgtgggt 993
Db 302 ACCGATTTCATTCTACGACGGGTGCACCTGCATAATGTGTGGTGGAAAGGNGGGG 361
QY 994 gaggacgcgattaccgttaagccaaacagcgcgggcaaaaatcccacgttgaaatcact 1053
Db 362 AAAACGCGGCCACTTCTCCGGAGCACCAGCGGGTGCAGGAGTACAGTTTCACGGTCA 421
QY 1054 aacagttcttcagacgcgctctgacaagaatctcagctgaatgcc 1101
Db 422 GCGCGCGGTGTGAAAATGCCAAGGACAAAGTGTTCAAAATGAACGCG 469

RESULT 2
LOCUS

CNS0060N 910 bp DNA GSS 03-JUN-1999
Drosophila melanogaster genome survey sequence T7 end of BAC #
BACR14J21 of RPCI-98 library from Drosophila melanogaster (fruit

fly), genomic survey sequence.	RESULT 3	CNS008MZ	959 bp	DNA	GSS	03-JUN-1999
AL065629	CNS008MZ					
AL065629.1	LOCUS					
GSS.	DEFINITION					
fruit fly.						
Drosophila melanogaster						
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;						
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;						
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.						
1 (bases 1 to 910)						
Genoscope.						
Direct Submission						
Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage						
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr						
- Web : www.genoscope.cns.fr)						
Determination of this BAC-end sequence was carried out as part of a						
collaboration with the Berkeley Drosophila Genome Project (BDGP).						
The BDGP is constructing a physical map of the Drosophila						
melanogaster genome using these BACs. For further information						
please see http://www.fruitfly.org The BDGP Drosophila						
melanogaster BAC library was prepared by Kazutoyo Osoegawa and						
Aaron Mammmer in Pieter de Jong's laboratory in the Department of						
Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,						
NY. The library is named RPCI-98 and was constructed by partial						
EcoRI digestion of Drosophila DNA provided by the BDGP from the						
isogenic strain y2; cn bw sp, the same strain used for the BDGP's						
p1 and EST libraries. A more detailed description of the library						
and how to order individual BAC clones, the entire library, or						
filters for hybridization from the BACPAC Resource Center can be						
found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.						
Location/Qualifiers						
1. .910						
/organism="Drosophila melanogaster"						
/db_xref="taxon:7227"						
/clone_lib="RPCI-98"						
/clone="BACR14J21"						
/note="end : 77"						
BASE COUNT 202 a 63 c 112 g 198 t 335 others						
ORIGIN						
Query Match 3.5%; Score 47; DB 219; Length 910;						
Best Local Similarity 19.3%; Pred. No. 0.02; 179; Indels 0; Gaps 0;						
Matches 78; Conservative 148; Mismatches 179; Indels 0; Gaps 0;						
QY 221 accagactacagagtggtgaacgctggcgctgaacgagcagaaagccacagcaggaa 280						
: : : : : : : : : : : : : : : : : : : : : : : : :						
Db 869 RSVCSACGAASAGASASASSAVSACGSSSSSVSSVSSASASAGVSRSSVSV 810						
QY 281 ccaactccgagctgcagctcagacacatcgtgagtgagatggcgcaacacgctgac 340						
: : : : : : : : : : : : : : : : : : : : : : : : :						
Db 809 SSVGSASRSASVSSSSSAGASAVAMACAMVASAVASVMSAVVAARASAKAAS 750						
QY 341 aggcacacacgccatgctccagcggcgcgagcagatcgcgataactcttactgaag 400						
: : : : : : : : : : : : : : : : : : : : : : : : :						
Db 749 MARVAAVARVASAAVVASASAAVAAVSRVSASARVMSASSSSSSSSSASAS 690						
QY 401 ccaatgctgaagttattgcacgcagtatgagcgccgaagcagatcagtttgccaaactg 460						
: : : : : : : : : : : : : : : : : : : : : : : : :						
Db 689 SSASSASMVVCRSAVSASAAASASASASASASASASASASASASASASAS 630						
QY 461 gtacggcgaacacacgctcttcggtactcttcttccttcggttccttccttcaag 520						
: : : : : : : : : : : : : : : : : : : : : : : : :						
Db 629 SSRCCGASSSSASASGSSSSSSSSCGCGSASASASASASASASASASAS 570						
QY 521 atctatcaggggggagggcccttcgcgaactcccttcgcggaactactcttcctcg 580						
: : : : : : : : : : : : : : : : : : : : : : : : :						
Db 569 GSS 510						
QY 581 gtacctttcaccaccatcagcgaacagtgctcccttaactcaccgc 625						
: : : : : : : : : : : : : : : : : : : : : : : : :						
Db 509 CASMCAMAAHACCCSCCCSCCCSCCCSCCCSCCCSCCCSCCCSCCCSCCC 465						
fly), genomic survey sequence.	RESULT 4	CNS0052P	844 bp	DNA	GSS	03-JUN-1999
AL056652	CNS0052P/c					
AL056652.1	LOCUS					
GSS.	DEFINITION					
fruit fly.						
Drosophila melanogaster						
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;						
Pterygota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;						
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.						
1 (bases 1 to 910)						
Genoscope.						
Direct Submission						
Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage						
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr						
- Web : www.genoscope.c						


```

Db 312 CCTCTTCAGCTCCTCTCTCCAGCTCCTCTCTCCATATTCAGCCTCTT 371
Qy 599 ccagccaagctccctaccctaccggttgatttcctttctccacc 648
Db 372 CGACCTCTTCTCTCTCTCCATCTTCGACCTCTCTCTCTATCTTCGACC 421

RESULT 13
AW173078
LOCUS
DEFINITION
XJ82H09.x1 Soares_NFL_T_GBC_S1 Homo sapiens cDNA clone
IMAGE:2663777 3' similar to contains element MSRI repetitive
element ;, mRNA sequence.
ACCESSION
AW173078
VERSION
AW173078.1 GI:6439026
KEYWORDS
EST.
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
1 (bases 1 to 534)
AUTHORS
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL
Unpublished (1997)
COMMENT
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
This clone is available royalty-free through LLNL ; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Seq primer: -40UP from Gibco
High quality sequence stop: 466.

FEATURES
source
1..534
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2663777"
/clone_lib="Soares_NFL_T_GBC_S1"
/lab_host="DH10B"
/notes="Organ: pooled; Vector: pT7T3D-Pac (Pharmacia) with
a modified polylinker; Site_1: Not I; Site_2: Eco RI;
Equal amounts of plasmid DNA from three normalized
libraries (fetal lung NBHL19W, testis NHT, and B-cell
NCI-CGAP_GCB1) were mixed, and ss circles were made in
vitro. Following HAP purification, this DNA was used as
tracer in a subtractive hybridization reaction. The driver
was PCR-amplified cDNAs from pools of 5,000 clones made
from the same 3 libraries. The pools consisted of
I.M.A.G.E. clones 297480-302087, 682632-687239,
726408-728711, and 729096-731399. Subtraction by Bento
Soares and M. Fatima Bonaldo."
BASE COUNT 102 a 174 c 68 g 190 t
ORIGIN

Query Match 3.1%; Score 41.2; DB 112; Length 534;
Best Local Similarity 60.9%; Pred. No. 0.79;
Matches 67; Conservative 0; Mismatches 43; Indels 0; Gaps 0;

Qy 539 ccccttcggcaactcccttcggcaactactctccgctgaagtacattctccaccat 598
Db 312 CCTCTTCAGCTCCTCTCTCCAGCTCCTCTCTCCATATTCAGCCTCTT 371
Qy 599 ccagccaagctccctaccctaccggttgatttcctttctccacc 648
Db 372 CGACCTCTTCTCTCTCTCCATCTTCGACCTCTCTCTCTATCTTCGACC 421

RESULT 14
AI336864
LOCUS
DEFINITION
qx86h08.x1 NCI-CGAP_GC6 Homo sapiens cDNA clone IMAGE:2009439 3'
similar to contains element MSRI repetitive element ;, mRNA

```

```

sequence.
AI336864
VERSION
AI336864.1 GI:4073791
KEYWORDS
EST.
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
1 (bases 1 to 563)
AUTHORS
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL
Unpublished (1997)
COMMENT
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Christopher A. Moskaluk, M.D., Ph.D., Michael
R. Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
Bonaldo, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html
Insert Length: 1179 Std Error: 0.00
Seq primer: -40UP from Gibco
High quality sequence stop: 448.

FEATURES
Location/Qualifiers
source
1..563
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2009439"
/clone_lib="NCI-CGAP_GC6"
/tissue_type="pooled germ cell tumors"
/lab_host="DH10B"
/notes="Vector: pT7T3D-Pac (Pharmacia) with a modified
polylinker; Plasmid DNA from the normalized library
NCI-CGAP_GC4 was prepared, and ss circles were made in
vitro. Following HAP purification, this DNA was used as
tracer in a subtractive hybridization reaction. The driver
was PCR-amplified cDNAs from a pool of 5,000 clones made
from the same library (clones 1257096-1258631,
1469064-1470983, and 1475592-1476743). Subtraction by
Bento Soares and M. Fatima Bonaldo."
BASE COUNT 113 a 176 c 67 g 207 t
ORIGIN

Query Match 3.1%; Score 41.2; DB 19; Length 563;
Best Local Similarity 60.9%; Pred. No. 0.8;
Matches 67; Conservative 0; Mismatches 43; Indels 0; Gaps 0;

Qy 539 ccccttcggcaactcccttcggcaactactctccgctgaagtacattctccaccat 598
Db 354 CCTCTTCAGCTCCTCTCTCCAGCTCCTCTCTCCATATTCAGCCTCTT 413
Qy 599 ccagccaagctccctaccctaccggttgatttcctttctccacc 648
Db 414 CGACCTCTTCTCTCTCTCCATCTTCGACCTCTCTCTCTATCTTCGACC 463

RESULT 15
AL566523/c
LOCUS
DEFINITION
AL566523 LTI_FL013_FBrnl Homo sapiens cDNA clone CS0DF020YA13 3
prime, mRNA sequence.
ACCESSION
AL566523
VERSION
AL566523.1 GI:12918965
KEYWORDS
EST.
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

```

Search completed: October 21, 2001, 11:30:56
Job time: 2739 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 21, 2001, 10:17:57 ; Search time 1309.91 Seconds
(without alignments)
15870.293 Million cell updates/sec

Title: US-09-596-958-1
Perfect score: 1344
Sequence: 1 atgtcaattcttaacgcttaa.....acctgaaggtgctgaatga 1344

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1344157 seqs, 7733874588 residues

Total number of hits satisfying chosen parameters: 2688314

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

GenEmbl:*
1: gb_ba1:*
2: gb_ba2:*
3: gb_ba3:*
4: gb_in1:*
5: gb_in2:*
6: gb_in3:*
7: gb_om:*
8: gb_ov:*
9: gb_pat1:*
10: gb_pat2:*
11: gb_ph:*
12: gb_pli:*
13: gb_pl2:*
14: gb_pl3:*
15: gb_pl4:*
16: em_ba1:*
17: em_ba2:*
18: em_fun:*
19: em_htgo_hum:*
20: em_htgo_inv:*
21: em_htgo_rod:*
22: em_htg_hum1:*
23: em_htg_hum2:*
24: em_htg_hum3:*
25: em_htg_hum4:*
26: em_htg_hum5:*
27: em_htg_hum6:*
28: em_htg_hum7:*
29: em_htg_hum8:*
30: em_htg_inv1:*
31: em_htg_inv2:*
32: em_htg_other:*
33: em_htg_rod:*
34: em_hum1:*
35: em_hum2:*
36: em_hum3:*
37: em_hum4:*
38: em_hum5:*
39: em_hum6:*
40: em_hum7:*
41: em_in:*
42: em_om:*
43: em_or:*

44: em_ov:*
45: em_pat:*
46: em_ph:*
47: em_pl:*
48: em_ro:*
49: em_sts:*
50: em_sy:*
51: em_un:*
52: em_v1:*
53: gb_sts1:*
54: gb_sts2:*
55: gb_sts3:*
56: gb_sy:*
57: gb_un:*
58: gb_v11:*
59: gb_v12:*
60: gb_htg1:*
61: gb_htg2:*
62: gb_htg3:*
63: gb_htg4:*
64: gb_htg5:*
65: gb_htg6:*
66: gb_htg7:*
67: gb_htg8:*
68: gb_htg9:*
69: gb_htg10:*
70: gb_htg11:*
71: gb_htg12:*
72: gb_htg13:*
73: gb_htg14:*
74: gb_htg15:*
75: gb_htg16:*
76: gb_htg17:*
77: gb_htg18:*
78: gb_htg19:*
79: gb_htg20:*
80: gb_htg21:*
81: gb_htg22:*
82: gb_htg23:*
83: gb_htg24:*
84: gb_htg25:*
85: gb_pr1:*
86: gb_pr2:*
87: gb_pr3:*
88: gb_pr4:*
89: gb_pr5:*
90: gb_pr6:*
91: gb_pr7:*
92: gb_pr8:*
93: gb_pr9:*
94: gb_rol:*
95: gb_ro2:*
96: gb_in4:*
97: gb_pri0:*
98: em_ba3:*

pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result	No.	Score	Match	Length	DB	ID	Description
	1	1344	100.0	2708	3	U94513	U94513 Erwinia amy
C	2	1344	100.0	9985	2	EADSPAB	Y13831 Erwinia amy
	3	1329.6	98.9	4291	1	AF083620	AF083620 Erwinia a
C	4	414	30.8	8143	2	EAU97504	U97504 Erwinia amy
	5	177	13.2	1833	1	AF037983	AF037983 Pseudomon
	6	155.2	11.5	1729	1	AF005221	AF005221 Pseudomon
	7	155.2	11.5	20052	16	AF232006	AF232006 Pseudomon
	8	155.2	11.5	52498	2	AF232004	AF232004 Pseudomon

9	95.6	7.1	1245	1	AB011839
c 10	91	6.8	47739	1	AF017113 Bacillus
11	91	6.8	209510	2	BSU00018
12	63.2	4.7	1214	2	BSP237980
13	62.2	4.6	884	14	FSOPELAB
14	61.4	4.6	40549	3	SCZG38
15	50.2	3.7	1419	14	FSU13050
16	45.8	3.4	7218	10	I66494
17	43.4	3.2	1332	14	FSU13049
c 18	43.2	3.2	7218	10	I66494
19	42.8	3.2	95432	14	ATT9E8
20	42.8	3.2	199634	13	ATCHRIV36
21	42.4	3.2	2185	14	FSOPELAA
22	42.4	3.2	213685	83	AP002760
c 23	42.2	3.1	209119	76	AC083818
c 24	42	3.1	237678	94	AC055818
c 25	41.8	3.1	152528	64	AC016063
c 26	41.8	3.1	183099	69	AC05288
c 27	41.6	3.1	5664	1	AE005748
c 28	41.6	3.1	68326	72	AC055793
c 29	41.2	3.1	2805	7	AF286474
c 30	41.2	3.1	82024	67	AC023210
c 31	41.2	3.1	194834	75	AC078873
c 32	40.8	3.0	203587	60	AC007491
c 33	40.6	3.0	221840	63	AC012618
c 34	40.2	3.0	169434	90	AL359205
c 35	40.2	3.0	171588	67	AC023198
c 36	40.2	3.0	198692	78	AC090530
c 37	40	3.0	135039	86	AC06060
-38	39.8	3.0	156975	81	AL442123
c 39	39.8	3.0	170396	81	AL512424
c 40	39.8	3.0	192730	97	HU095743
c 41	39.4	2.9	88036	97	HUAC003119
c 42	39.2	2.9	162342	80	AL359966
-43	39	2.9	166735	79	AL355861
c 44	38.8	2.9	861	14	CRCCGR4
c 45	38.8	2.9	90832	85	AC003065

ALIGNMENTS

RESULT	1
U94513	U94513 2708 bp DNA BCT 03-OCT-1998
LOCUS	Erwinia amylovora disease-specific operon, partial sequence; and
DEFINITION	hrpW operon, complete sequence.
ACCESSION	U94513
VERSION	U94513.1 GI:3414585
KEYWORDS	type III protein secretion; Hrp pathway; hypersensitive response; peptate lyase.
SOURCE	Erwinia amylovora.
ORGANISM	Erwinia amylovora.
REFERENCE	Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae; Erwinia.
AUTHORS	Kim, J.F., Zumoff, C.H. and Beer, S.V.
TITLE	HrpW, a new harpin of Erwinia amylovora, is a member of a family of peptate lyases
JOURNAL	Phytopathology 87, 552 (1997)
REFERENCE	2 (bases 1 to 2708)
AUTHORS	Kim, J.F. and Beer, S.V.
TITLE	HrpW of Erwinia amylovora, a new harpin that contains a domain homologous to peptate lyases of a distinct class
JOURNAL	J. Bacteriol. 180 (19), 5203-5210 (1998)
MEDLINE	98422475
REFERENCE	3 (bases 1 to 2708)
AUTHORS	Kim, J.F. and Beer, S.V.
TITLE	Direct Submission
JOURNAL	Submitted (18-MAR-1997) Plant Pathology, Cornell University, 334 Plant Science Bldg., Ithaca, NY 14853, USA
FEATURES	Location/Qualifiers
source	1. .2708

/organism="Erwinia amylovora"	
/strain="Ba321; ATCC 49947"	
/db_xref="taxon:552"	
/db_xref="ATCC:49947"	
/note="a strain that infects apples, pears, and other pomaceous plants"	
/complement(<1. .784)	
/gene="disease-specific operon"	gene
/complement(<1. .729)	gene
/gene="dspe"	CDS
/complement(<1. .714)	
/gene="dspe"	
/note="Hrp-secreted pathogenicity/avirulence protein similar to Pseudomonas syringae AvrE"	
/codon_start=1	
/transl_table=11	
/product="dspe"	
/protein_id="AAC62315.1"	
/db_xref="GI:3414587"	
/translation="MELKSLGTEHKAHVTAHNPVGHVVALQOQSSSSPQNAASL AEGKNGKMPRIHOPSTAAADGISAHQKSFSLRGLGTFKFSRAPOGQPGTHS KGATRLDLPARDGCTOHEAAPDAARLTBSGGVKKRNMDDMAGRPVKGSGDKVP TQKRHLQNFQGMQRQWLSKMAHPASANAGDRLQHSPPHPCSHHEIKKEEPVGSTK ATTAAHARVEIAQEDDDSEF"	
/complement(724. .729)	RBS
/gene="dspe"	
/complement(755. .784)	promoter
/gene="disease-specific operon"	
/note="HrpL-dependent promoter consensus sequence; hrp box"	
825. .854	promoter
/gene="hrpW operon"	
/note="HrpL-dependent promoter consensus sequence; hrp box"	
825. .2692	gene
/gene="hrpW operon"	RBS
900. .905	
/gene="hrpW"	gene
900. .2255	
/gene="hrpW"	
/note="harpin gene"	CDS
912. .2255	
/gene="hrpW"	
/function="elicit the hypersensitive reaction"	
/note="Hrp-secreted; C-terminal domain similar to class III peptate lyases; high glycine and serine content; lacks cysteine"	
/codon_start=1	
/transl_table=11	
/product="HrpW"	
/protein_id="AAC62314.1"	
/db_xref="GI:3414586"	
/translation="MSILTUNNTSSPGLFQSGGDNGLGHNANSALGQOPTDQRTI EQMAQLIAELLKSLSPSGNAATGAGNDQTTGVNAGGLNGRGTAGTTPQSDSON MLSBMNGNGLDQAITPDGQGGQIGDPLKMLKLIARMMDGSDGFGQPGGNNSA SSGTSSSGSGPFDNLGSGKAPSGNSPGNSPVSTFSPSTPTSPDLDPSPPTKA AGGSTPTVDPDPVGSAGIGAGNSVFTSAGANQTVLHRTITVKAGOVFDGKGTFTA GSELGDGQSGENKQPLIEDGASLKNVTMGDDGADGCIHLGDKAKINLHVTNVEDA ITVPNSAGKSHVEITNSSEFHASDKILQNLADTNLSVDNVRKADQFTGRTNGGOO GNMDNLNLSHISAEDGKFSFKVKSDEGLNVNTSDISLGDVENHYKVPMSANLKVAE"	
2242. .2247	RBS
2252. .2584	CDS
/note="putative; OrfC; high glutamine content"	
/codon_start=1	
/transl_table=11	
/product="potential HrpW-specific chaperone"	
/protein_id="AAC62316.1"	
/db_xref="GI:3414588"	
/translation="MMKSAEIPPAESWVDVNSQOHLSPAQYQAFQKAIQDOVORLH QVLSRPMPOQKGKFDLSFVDSLHADFLIGSDSDSDSNVGTQANWITRYLADRLLELQ QLOQQRGL"	
2655. .2692	terminator
/gene="hrpW operon"	

RESULT 2

ACCESSION

**RECORDS
SOURCE**

AUTHORS

△

MEDLINE
PREFERENCE

TITLE

REMARK

AUTHORS

FLUORON

[illegible]

MEDLINE

FEATURES
SOURCE

```
/clone="pMAB37"
/lab_host="pear and apple trees"
/complement(1605..3035)
/gene="hrpW"
/cds
  complement(1605..2948)
/gene="hrpW"
/codon_start=1
/transl_table=11
/product="HrpW protein"
/protein_id="CAA74158.1"
/db_xref="GI:3150150"
/db_xref="SPTREMBL:O54508"
/translation="MSIITLNNNTSSSPCLFQSGDNGLGGHNSALGQQPIDRQTI
EQMAQLLESLKSLSPGAGNAGTQGTGAGNAGLNGRGTAGTTPQSDSON
MLSENGNGLDQAIPTDQGGQIGDNPDLKAMLLIARMMDGSDQFQCPQTGNNSA
SSGTSSSGSPNDLSGGKAPSGNSPVSPSPSPSPSPSPSPSPSPSPSPSPSPSP
AGGSTPVTDHPDPVGSAGIGAGNSVAFTSAGANQTVLHDITVRAGQVDFKGTQFTA
GSELGDGSGENRPLFLIEDGASLKNVTMGDDGADGHLHYKADKIDMLNVTGEDA
ITVPNSAGKXSHVETITNSFEHSDKTLQLNADTNLSDVNYKADFCTFVRTNGGQ
GNMDLNLSHISAEDGKFFSVKSDSEGLNVNTSDISLGDVENHYKVPMSANLKVAE"
/complement(3008..3035)
/gene="hrpW"
/promoter
  3076..3101
/gene="dsbA"
/gene="dsbA"
  3076..8662
/gene="dsbA"
  3146..8662
/gene="dsbA"
/function="involved in pathogenicity"
/codon_start=1
/transl_table=11
/product="DspA protein"
/protein_id="CAA74156.1"
/db_xref="GI:2695858"
/db_xref="SPTREMBL:O54581"
/translation="MELKSLGTEHRAAVHTAAHNPVGHGVALQGGSSSSPQNAASL
AAEGRNKMRIPIHOPSTADGISAHQOKFSFLRGCLGTFKFSRSPQPOGTTHS
KGATLRDLNDRDQTEHEAAPAARLTRSGGVKRRNMDMDMAGRPVKGSGEDKVP
TQQRHLQNNQCOMQRTMLSKMAHPASANAGDRLOHSPPHIPGSHHEIKESPVGSTK
ATTAHADRETAQEDDSEFOQLHOORLARENPPOPKLGVAFTPI SAROPKLTAV
AESVLEGTDTTQSPKPSMLKSGAGVTPLAVTLDKGLQLADPNPDKLTKQTL
KQDQHYLAHASDSCGSHLLDNKGHLFDIKSTATSYSLNHSHPGKIKLQAQTL
GVSVDGKSLGSGTQSHNKTMLSPQGEAHRSLTGIWOHPAGAARPOGESIRLH
DQKHILHPELQWSADKTHSOLSRQADGKLYALKDNRTLONLSDNKSSEKLVDKI
KSYVDQGRVAILTDTPGRHKMSIMPISLDASPEHSISLHFADAHQGLHQSSELE
AQSAVSHGRVAVDSEKLFSAALPKQGDGNELKMKAMPQHALDEHGHDOHSQFF
HDDGQALNALVKNPFRQOCHPLGNDHOFHPGWNLTDAVLIDNOLGLHHTPEPHEIL
DMHGLSLALQEGKLIHYFDQLTKWGTAESECKQLKGLDGAAYLLKDKCEYKRLNIG
STSLKHTGENFSLPHVRNRPEDGALQGLNKKDKAAMAVIGYNNKYLALETGKDIR
SFQIKPGTQQLERPAQTLREGISGELDKIDHVKHQNLYALTGHEGEVHPQREAWQNG
VGTSGQTVFNRMLMGVKGKVI PGSLTVKLISAOTGMGTGAERKVSXKFSERTRAYA
FNPTMSTPRPKNAAYATQHGQWQREGKLPYEMOGALIKOLDAHNVHRNAPQPDLOS
KLETLDLGEHGAELLNDMKRPRDELQOASTVSVTLGQHGVLKSNCEINSEFPSPG
KALVSPNNSGQDLSLQQAQVHATPPSAESKLQSLMGLHFVSAGVDMSHQKGEIPL
GRQDPNDKTLASRLLDVTITGELHELADKALVSDHDPDADQIKOLROQFDTLR
EKRESNPVKHYTDGTFHNKALEANYDAVKAFTNAFKKEHGVNLTTRTVLESQGS
ELAKKLNTLLSLDSEMSFSRSYGGVSVFVFTLSKKVPVPIPGAGITLDRAYN
LSFSTSGGLNVSGFRDGSVGNIMVATGHVMPYMTCKTTSAGNASDMLSAKHKISP
DLRTICAASGTGLQNSLKEFLTEDELPFIHGLTHGTLTPAELLQKGEHOMKQK
SLRTFSVDTSLNDLURAGINLNEGSKPNGVARTVSAGLSANLAAGRSRSTTSGQ
FGSTTSANNRPRTFLNGVAGANLTAALGAHSTHCKPVGIPFAPFTSTNVSAALAL
DNRTSQSLSLKLEAPYTSNDISELTSLGKHFKDSATTKMLAAKELDDAKAEQL
HILQOHFSKADKVVGDVRYVARNLKLVIROQAADSHSMLGSASHSTYNNLRINN
DGIIVELLKHFDAAALPASAKRLGEMMNDPALDKI IKQLOSTPSSASVSMELKDLG
REOTEKATILDGKVEEYGVLFQDRNNLRVKSVSQSVSKSEGFNTALLGTSNSA
AMSERNILGKINFTGQQNPRTFLEGGIAQANPQVASALTDLKKEGLEMK"
  8675..8701
/gene="dsbB"
  8675..9143
/gene="dsbB"
  8724..9143
/gene="dsbB"
/promoter
```

```
/function="involved in pathogenicity"
/codon_start=1
/transl_table=11
/product="DspB protein"
/protein_id="CAA74157.1"
/db_xref="GI:2695859"
/db_xref="SPTREMBL:O54621"
/translation="MTSSQQRVERFLQYFSAGCKTPIHLKDGVCALYNEQDEEAAVLE
VPOHSDLLHCRITIEADPTSTIYSLMLLOLNFEAMARCMGLALDELHNVRLCFQO
SLEHLDEASFSDIVSGFTEHAAREVYIAQLEDESSAA"
BASE COUNT 2570 a 2562 c 2726 g 2127 t
ORIGIN
Query Match 100.0%; Score 1344; DB 2; Length 9985;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1344; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 atgtcaattcttacgtcttaacaacaataacctctctctcgcgggtctgttccagtcggg 60
Db 2948 ATGTCAATTTCTTACGCTTAACAACAATACCTCGTCTCTCGCGGGTCTGTTCAGTCCGGG 2889
QY 61 ggggacaacggcttgggtggtcataatgcaaatcttgggtggggaacacccatcgat 120
Db 2888 GGGGACAACGGGCTTGGTGTGTCATAATGCAAAATCTGCGTTGGGGCAACAACCCATCGAT 2829
QY 121 cggcaaacattgagcaaatggctcaattattggtgggaactgttaagtcactgctatcg 180
Db 2828 CGGCAAAACCATTTGAGCAAAATGGCTCAATTTATTGGCGGAACCTGTTAAAGTCATGCTATCG 2769
QY 181 ccacaatcaggtaattgctgggcaacccggagccggtggtgcaatgaccagactacagagtggt 240
Db 2768 CCACAATCAGGTAATGCGGCAACCGGAGCGGTGCAATGACCACTACACGAGTGTGAT 2709
QY 241 aacgtggtggcctgaacggagcaaaagagcagcaggaacacactccgcagtcctacagct 300
Db 2708 AACGTGGCGGCTTGAACGGAGCAAAAGAGCACAGCAGAGGAACCACTCCGCGAGTCTGACAGT 2649
QY 301 cagaacatgctgagtgagatgggcaaacggcggtggatcagggccatcacgcccagatggc 360
Db 2648 CAGACATGCTCAGTGAGATGGGCAACACGCGGTGGATCAGGCCATCAGCCCCGATGGC 2589
QY 361 caggcggtggcagatggcggaataatccttctactgaaagccatgctgaagcttattgca 420
Db 2588 CAGGCGGCGGCGAGATCGCGGATAATCTTTACTTGAAGACCATGCTGAAGCTTATTGCA 2529
QY 421 cgcatagtgacggccaaagcagatcagtttggcaacacctggtacgggcaacacagtgcc 480
Db 2528 CGCATGATGGACGGCCAAAGCCGATCAGTTTGGCCAACTTGGTACGGGCAACACAGTGCC 2469
QY 481 tcttcgggtacttcttcactgctgggttccctctttaaagcactctatcagggggggaagggc 540
Db 2468 TCTTCCGCTACTTCTTCATCTCGCGGTTCCTCCCTTTTAAAGGATCTATCAGGGGGGAAGGCC 2409
QY 541 ccttcgggcaactccctcttcgggaactactctctccgctcagtcacccatccatcc 600
Db 2408 CCTTCCGGCAACTCCCTCTCCGGCAACTACTCTCCCGTCAGTACCTTCTTCACCCCATCC 2349
QY 601 acgcaaacgtccctctacactcaccgcttgatttccctcttctcccaaacacagccggg 660
Db 2348 AGCCCAAGCTCCCTTACCTTACCGCTTGATTTCCTTCTTCTCCCAACCAACAGACGCCGG 2289
QY 661 ggcagacgcccggtaaccgatcatctgacacctgttgtagacctgttgtagcggggatcgggccgga 720
Db 2288 GGCAGACGCCGGTAACCGATCATCTTGACCTCTTGGTAGCGGGCATCGGGCCGGA 2229
QY 721 aattcgggtgcttcacacgacgcccggcgctaatacagaoggtgtgtcatgacaccattacc 780
Db 2228 AATTGGGTGGCTTACCAGCGCGCGGCTAATAAGACGGTGTGCTGATGACACCATTTACC 2169
QY 781 gtagaagcgggtcaggtgtgttagtgcaagagcaaaccttccaccggtgttcagaatta 840
Db 2168 GTGAAGCGGGTCAAGTGTGTGTTGCGAAGGACAAACCTTCACCGCGGTTCAAGATTGA 2109
```

```
QY 841 ggcatggggccagctctgaaccagaaacccgctgtttatatactggaagcaggtgcagc 900
|||||
Db 2108 GGGCATGGCGCCAGCTCTGAACACCAAGAACCGCTGTTATATCTGAAGACGGTGCAGC 2049
QY 901 ctgaataacgtccaccatggcgacgacgagggcggaatcggtattcatctttacagtgatgcc 960
|||||
Db 2048 CTGAATAACGCTACCACTGGCGACGACGCGGGCGGATGCTTATCATCTTTACGGTGATGCC 1989
QY 961 aaaaatagacaactgcacatccacacagctgggtgaagacgagcattaccgttaagcacaac 1020
|||||
Db 1988 AAATAAGACAATCTGCAGCTCAACACGCTGGGTGAGACGCGATTACCGTTAAGCCAAAC 1929
QY 1021 agcgggggcaaaaaatccacagcttgaaatcaactaacagttcccttcagcacgcctctgac 1080
|||||
Db 1928 AGCGCGGCAAAAATCCACGCTGGAATCACTAATCACTTCTTCGACGACGCTCTGAC 1869
QY 1081 aagatcctgcagtgtaatcccgatatacctacacagtgagcgttgacaacatgaagcgcaaac 1140
|||||
Db 1868 AAGATCTCGAGCTGAATCCGATACCTAACCTGAGCGTTGACAACTGAAGGCGCAAGAC 1809
QY 1141 ttctgactttgtacgactaacgcgctcaacaggttaactggaactggaatctgaac 1200
|||||
Db 1808 TTTGGTACTTTGTACGCACTAACGCGGTCAACAGGGTAACCTGGGATCTGAATCTGAGC 1749
QY 1201 catatcagcgcaagaacacggttaagttctctgtcttcttaaaagcgatagcgagggctaaac 1260
|||||
Db 1748 CATATCAGCGCACAGACGGTAGTTCTGCTGCTTTAAAGCGATAGCGAGGGGCTAAAC 1689
QY 1261 gtcataacagtgatctcactcgggtgtagtgtgaaacacactacaaagtgcgatgtcc 1320
|||||
Db 1688 GTCAATACCACTGATATCTCACTGGGTGATGTTGAAACCACTACAAAGTGCCGATGTCC 1629
QY 1321 gccaaactaaagtgcctaataa 1344
|||||
Db 1628 GCCAACTGAAGTGCTGAATGA 1605
```

RESULT 3

```
AF083620/c
LOCUS
DEFINITION
Erwinia amylovora harpin HrpN (hrpN) gene, partial cds; potential
ORFB-specific chaperone, virulence/avirulence effector protein
homolog, probable HrpW-specific chaperone, and harpin HrpW genes,
complete cds; and Hrp-secreted pathogenicity/avirulence protein
DspE (dspE) gene, partial cds.
ACCESSION
AF083620
VERSION
AF083620.1 GI:7542322
KEYWORDS
Erwinia amylovora.
SOURCE
Erwinia amylovora
ORGANISM
Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
Erwinia.
REFERENCE
1 (bases 1 to 4291)
AUTHORS
Kim,J.F., Laby,R.J. and Beer,S.V.
TITLE
Comparison of the hrpN-flanking regions of two Erwinia amylovora
strains with different host specificity
JOURNAL
Unpublished
REFERENCE
2 (bases 1 to 4291)
AUTHORS
Kim,J.F., Zumoff,C.H. and Beer,S.V.
TITLE
Direct Submission
JOURNAL
Submitted (12-AUG-1998) Plant Pathology, Cornell University, 334
Plant Science Bldg., Ithaca, NY 14853, USA
FEATURES
source
1. .4291
/organism="Erwinia amylovora"
/strain="Ea246"
/db_xref="taxon:552"
/note="pathogenic to Rubus plants"
<1..429
/gene="hrpN"
CDS
<1..429
/gene="hrpN"
```

```
/note="HrpN family harpin"
/codon_start=1
/transl_table=11
/product="harpin HrpN"
/protein_id="AAF63398.1"
/db_xref="GI:7542323"
/translation="GTGIGMKAGIQALNDIGTHSDSSTRSFVNKGDRAKAIQGFMD
QYVEFGKPYQKQPGQEVKTDKSWAKALSPPDDGTMTPASMEQFNKAKGIKSAMA
GDTGNGNLQARGAGSSSLGIDAMMAGDAINNNAKLGAA"
472..477
RBS
481..491
CDS
/note="ORFA; member of Sych family of chaperones and
chaperone-like proteins; leucine-rich"
/codon_start=1
/transl_table=11
/product="potential ORFB-specific chaperone"
/protein_id="AAF63399.1"
/db_xref="GI:7542324"
/translation="NMIIQDLGALARRLEAGTSLSEAGQLCCQLVNGLEMTLEWLEQ
QNMFLVLSVGTLSGTSTLLSILAANLPHYGSDGAAGLDRKNELELLFORFQL
SSVDEASPVSAQVQMEIAVKRWQKFLHGRALPTAAPRLAQPLGLTLNLAGIK"
936..941
RBS
948..2129
CDS
/note="ORFB; similar to ORFB of Erwinia amylovora Ea321,
AvrRxv of Xanthomonas campestris pv. vesicatoria, AvrA of
Salmonella enterica, and YopJ of Yersinia
pseudotuberculosis"
/codon_start=1
/transl_table=11
/product="virulence/avirulence effector protein homolog"
/protein_id="AAF63400.1"
/db_xref="GI:7542325"
/translation="NNISGLRGYKSOAQADNASSSSSTOGSPAPTGRRLRQDALPA
NYRYHSAOMPATPERVAARYASQASSAGPSILLRSQSHRENPRLVRFHETWQOS
PKMSRGDLPKPEIVPKRLQEKIDSYNLPRNLKDKNLYEKNATELAKGSGSSV
ALMRDKKVPLLLADAENARNPGLNHYVKGEECYQAIKEQHKIYQOSGQPKTRAL
YPPFMPDHHIALDIHLRPHRPSIVGFSALGHMVDPIROGIAOGLRGAKVHMVG
QIONSEWDCIMYSLNNAKLSFKHDEYTLARLHKGEKIPVPAEFFHAOSKSMVEGLPH
ODAIVTKDKGGLHAETLLHRLNLAIRADRFHACNTSIEGFRMQEIQRAEPLSAQNRK
S"
2145..2184
terminator
CDS
complement(2253..2582)
/note="ORFC"
/codon_start=1
/transl_table=11
/product="probable HrpW-specific chaperone"
/protein_id="AAF63401.1"
/db_xref="GI:7542326"
/translation="MMKSAEIPPAESWVDVYVNSQORLSPEQYQAFQRAISQVKORLQ
QVLGRMPQROGEFELDSFVLSHADFLNVDGKAANDVDVGQTAWITRYLADRLLELQ
QLEQQRG"
gene
complement(2579..3934)
/gene="hrpW"
CDS
complement(2579..3922)
/gene="hrpW"
/note="HrpW family harpin; C-terminal domain similar to
class III pectate lyases"
/codon_start=1
/transl_table=11
/product="harpin HrpW"
/protein_id="AAF63402.1"
/db_xref="GI:7542327"
/translation="MSILTLNSNTSSPGLFSGGGDNLGGHNANSALGQOPIQDRTI
EQMAQLLAELLSLSPQSNATGAGNDQDTGVGNAGGLNCRGTAGTGTTPQSDSON
MISEMNGLDQAITPDGOGGOIGDNPULKAWLKLIARMDDGQSDQFGQPGTGNSA
SSGTSSSGSGPNDLSSGGKAPSGNSYSPVSTFSPSTPTSPSTPSPSTPKA
AGGSTVTDHPDPVGSAGIGAGNSVFTSAGANTVLIHDTITVKAQSVDFDKGQFTFA
GSELGDGQSENOKPLFILEDGASLKNVTMGDGDAGDILHYDKADINDLHVTNGEDA
ITVKPNSAGKKSHEVITNSFEHASDKILQLNADTNLSVDNVKAKDFCTFVTNGGQQ
GNWDLNLSHISAEKSKFSPKSDSEGLKAVNTSDISLSDVENHYKVPMSANLKVAE"
2587..2592
RBS
complement(3929..3934)
/gene="hrpW"
RBS
```


source 1. .8143
/organism="Erwinia amylovora"
/strain="Ea321"
/db_xref="taxon:552"
/note="disease specific operon"
complement(<1. .501)
/gene="hrpW"
complement(<1. .414)
/gene="hrpW"
/note="harpin"
/codon_start=1
/transl_table=11
/product="HrpW"
/protein_id="AAC04849.1"
/db_xref="GI:2668587"
/translation="MSILTLNNNTSSPGLFOSGGDNGILGGNANSALGQQPIDROTI
EQMAQLLAELLSLSPQSGNAATGAGGNDQGTGVNAGLNRKGTAGTTPQSDSQN
MLSEMGNGLDQAIPTDGGGGQIGDNPILKAWLKL"
complement(473. .501)
/gene="hrpW"
box' "
/note="hrpL-dependent promoter consensus sequence, or 'hrp
box' "
542. .570
/gene="disease specific operon"
/note="hrpL-dependent promoter consensus sequence, or 'hrp
box' "
542. .6609
/gene="disease specific operon"
542. .6128
/gene="dspE"
612. .6128
/gene="dspE"
/note="AvrE homolog; pathogenicity factor; avirulence
protein"
/codon_start=1
/transl_table=11
/product="DspE"
/protein_id="AAC04850.1"
/db_xref="GI:2668588"
/translation="WELSLGTEHKAHVTAHNVPVGHCVALQQSSSSPONAASL
AEGNRKMPRIHQPSSTAADGISAHQKKFSFSLRGCLGTRKFSRSPQGGQTHS
KAGTRDLIARDGGTDEAAAPDAARLTSSGGVRRNMDMAGRPWYKGGSGEDVP
TQQRHLNQFMQMTMLSKMAHPANAGRLQHSPPHIPGSHHEKEEVPSTSK
ATTAAHRAVEIAQEDDEFEQHOURLARENPPPKLGVAATPISARFQPKLTAV
AESVLEGTDTTQSLPKPSMLKSGAGVTPLAATLDKGLQLAPNPALNTLLKQTL
GRDQHYLAHSSDGSQHLLDNKGHLEIDIKSTATSYSLNHPGEIKKLAQGT
GSVVDGSGKSLSGSGTQSHNKTMLSPQGEAHRSLTIGIWOHPAGAHPQSESLRLH
DDIHLPELQVWQADKTHSQRQADGLYALKDNRTLONLSDNKSSESLVDKI
KSYVDQRQVAILDTDFGRHKMSIMPSLDASPEHSISLHFAHAHQGLLHKSELE
AQSAVISHGLVADSEGLFSAATPKQGDGNELMKAMPQHALDFHGHDIHISGFF
HDDHQLNALVKNFRQOHACPLGNDHOFHPCWNLTDAVLIDNQLGHLHNPPEHIL
DMHGLSLALQSEKLYHFDOLTKGWTGAESDCKLKGLDGAAYLLKQGEVRELINQ
STSSIKHGTENVFSLPHVRNKEPQDGLQGLNKKDQKQAMVIGWYKYLALTEKGDIR
SFQIRPGTQQLERPAQTLSREGISGLDHIHVDHKNLYALTHEGEVHPQREAWNG
AESSWHKLLALPOSSEKLSLDMSEHKPIATFEDGSOHLKAGWHAYAAPERPLA
VGTSGVTNFRMLQGVKVIKPGSEGLTVKLSAQTGGMTEAGRKVSKFSERIRAYA
FNPTMTPRPIKNAAYATOHGQREGLEKPLVEMOGALIKOLDADHVRHNAPOPDLOS
KLETLDLGHGAELLNDMKFRFDELEQATRSVTVLGOHGVLKNGEINSEFKSPG
KALVQSFVNRSGDLSLQQAQVHATPPSAESKLQSMLEHGFVSAGVDMSHQKGEIPL
GQRDPDKTALTKRLIIDLTVTIGELHELADKALVSDHKPDADQIKOLRQOFDTLR
EKRYFENPVKHYTDGFTGHEKALEANYDAVKAFINAFKEHGVNLTTRVLESQSA
ELAKKLTLLSLDSGESMSFSRSGVSTVFVPTLSKKVPVPIPGAGITLDRAYN
LFSRTSGGLNYSFGRDGGVSGNINVATGHDVMPYMTGKTSAGNASDWSLAKHKISP
DLRIGAAVSGTQGLTQNSLSEKLTDELPFGTHGLTLPALLOKIEHOMKQ
SKLTSVSDNTSANDLRAGINLNEADSKPNVGTARVSAGLSANAASRSTSSGQ
FGSTYSASNRTPLFNGVAGANLTAALGVAHSSSTHEGKPVGIFPAFTSTVNSAALAL
DNRTSQSTLELKRPEVPTSDISLTSLTGLKFKDSDATTKMLAKLEDDAKPAEQL
HILOQHFSAQVDERVEAVRNKLVIROQAADSHSMELGSASHSTYNNLSRIN
DGIVELLHGHFPAALPASSAKRLGEMMNDPALKDIKLOLQSTPFSSASVSMELDGL
RQTEKAILDKGVGREEVGLFQDRNRLRVKSVSVSQSVSKSEGTNPALLLGTSNA
AMSMERNGTINFKYQQDQNTPRFTLEGGIAQANPQVASALTDLKKEGLEMKS"
6190. .6609
/gene="dspf"

CDS 6190. .6609
/gene="dspf"
/codon_start=1
/transl_table=11
/product="Dspf"
/protein_id="AAC04851.1"
/db_xref="GI:2668589"
/translation="MTSSOORVERFLQYFSAGCKTPILKDKGVCALYNBQDEAAVLE
VPOHSDSLLLHCRITTEADPQTSITLYSMILQINFMAAMRGCMALDELHLYRLCTFOQ
SLEHLDSEAFSDIVSGFIEHAAREVREYIAQLDESSAA"
7361. .>8143
/note="partial ORF"
/codon_start=1
/transl_table=11
/product="lysR homolog"
/protein_id="AAC04852.1"
/db_xref="GI:2668590"
/translation="MFISKMQVFTLYOEKSLKTAADKLCTVPPVSRMLKLTGEGV
GEQLFIERNITPTPAADCIYQQLPLLYMALKNMTRKHQEQEFLIASPYISTISLS
DLTITSEALPKCSTIKYAECHPHDDIYISLQPECPMHFPIRMDLVLECCSKDV
GDKLHMPTLAESETKLYKFORSLAELHARGSGVLRQVDNPSHLQSTFKSGEGLLF
RLPPKNMYKDYQSLPVISSHCPLELYVNTLKKDKRHNCFATL"
BASE COUNT 2181 a 2027 c 2177 g 1758 t
ORIGIN
Query Match 30.8%; Score 414; DB 2; Length 8143;
Best Local Similarity 100.0%; Pred. No. 7.1e-95;
Matches 414; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 atgtcaattcttaccgttaacaacaatacctcgtctcctcgcggggtctgttccagtcggg 60
Db 414 ATGTCATTTCTTACGCTTAAACAATACTCTGCTCGCGGGTCTGTTCAGTCCGG 355
QY 61 ggggcaacaggcgttggtggtcataatgcgaattctgcgttgggggcaacacccatcgat 120
Db 354 GGGGACACCGGGCTTGCTGTCATTAATCAAAATCTCGCTGGGGCAACACCCATCGAT 295
QY 121 cggcaacatttgagcaaatggctcaattatttgcggaactgttaagtcactgctatcg 180
Db 294 CGGCAACACCATTAAGCAAAATGGCTCAATATTGCGGGAAGCTGTTAAGTCACTGCTATCG 235
QY 181 ccacaatcaggtaatcggcgcaacccgagcggctggcaatgacagactacagaggttgg 240
Db 234 CCACATCAGTAATCGGCAACCCGCGGCTGGCAATGACCACTACAGGAGTGGT 175
QY 241 aacgctggcggtcgtgacgagcaagaaaggcagcagcaggaacactccgagtcgacagt 300
Db 174 AACGCTGGCGGCTGAACGAGCAAGAAAGGCACAGCAGGAACCACTCCGAGTCTGACAGT 115
QY 301 cagaacatgctgagtgagatgggcaacaacggcggtgagtcagcagcccgatggc 360
Db 114 CAGAACATGCTGAGTGAGATGGGCAACACGCGGTGGATCAGGCCATCAGCCCGATGGC 55
QY 361 caggcgcgcggtcagatcggcgataatcctttactgaaagccatgctgaagctt 414
Db 54 CAGCGCGCGGCGGAGATCGCGGATATCTTTACTGAAAGCCATCTGAGAGCTT 1
RESULT 5
AF037983
LOCUS AF037983 1833 bp DNA BCT 05-OCT-1998
DEFINITION Pseudomonas syringae hrp type III secreted protein (hrpW) gene,
complete cds.
ACCESSION AF037983
VERSION AF037983.1 GI:3695000
KEYWORDS
SOURCE Pseudomonas syringae pv. syringae.
ORGANISM Pseudomonas syringae pv. syringae
Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
Pseudomonas.
REFERENCE 1 (bases 1 to 1833)
AUTHORS Charkowski,A.O., Alfano,J.R., Preston,G., Yuan,J., He,S.Y. and

Qy	1108	aacctgagcgcttgacaaagtgaaagccaaagactttgtactttttgacgcactaaagcgc	1167
Db	1354	CACCTGAAAGTCGACGGGTTCACGAGCCACCGATTCCTGGTACGGTGTGTCGACCAATGGT	1413
Qy	1168	ggtaacacag--ggtaactaggatctgaatctagccatcatcagcgaagacggttaag	1224
Db	1414	GGCAGACGTTTGATGACATGAGTGTCTGAGCTCAATGGCGTGACGCGACCCATGGCAAG	1473
Qy	1225	ttctcgcttgctaaagcgcagatagcgagggggttaaacgtcaataaccagtgatatctcactg	1284
Db	1474	TTCCGACTCGTCFAAAACGACAGTGACAGATCTGAACACTGGCCACTGCGCAGACATCGCCATG	1533
Qy	1285	ggtagctgtgaagaaacacactacaagtgccgatgtccgcgaacactgaa	1331
Db	1534	ACGACGTCGACGCGCTTGATGACCAAGCAACCAAGGCGTCTACCCAGCA	1580

RESULT	6
AF005221	AF005221
LOCUS	1729 bp DNA BCT 05-OCT-1998
DEFINITION	Pseudomonas syringae hrp-secreted protein HrpW (hrpW) gene, complete cds.
ACCESSION	AF005221
VERSION	AF005221.1
KEYWORDS	GI:3694998
SOURCE	Pseudomonas syringae.
ORGANISM	Pseudomonas syringae
REFERENCE	Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae; Pseudomonas.
AUTHORS	1 (bases 1 to 1729) Charkowski,A.O., Alfano,J.R., Preston,G., Yuan,J., He,S.Y. and Collmer,A.
TITLE	The Pseudomonas syringae pv. tomato HrpW protein has domains similar to harpins and pectate lyases and can elicit the plant hypersensitive response and bind to pectate
JOURNAL	J. Bacteriol. 180 (19), 5211-5217 (1998)
MEDLINE	98422476
REFERENCE	2 (bases 1 to 1729) Charkowski,A.O., Conlin,A.K. and Collmer,A.
AUTHORS	Direct Submission
TITLE	Submitted (23-MAY-1997) Plant Pathology, Cornell University, Ithaca, NY 14853, USA
JOURNAL	Location/Qualifiers
FEATURES	1..1729
source	/organism="Pseudomonas syringae"
gene	/strain="DC3000"
promoter	/db_xref="taxon:317"
gene	/note="pathovar tomato"
promoter	pathovar: Lycopersicon esculentum"
gene	268..1693
promoter	/gene="hrpW"
gene	268..292
promoter	/gene="hrpW"
gene	/note="hrp box"
promoter	406..409
gene	/gene="hrpW"
promoter	419..1693
gene	/gene="hrpW"
promoter	/note="harpin; hrp-secreted protein"
gene	/codon_start=1
promoter	/transl_table=11
gene	/product="HrpW"
promoter	/protein_id="AAC62526.1"
gene	/db_xref="GI:3694998"
promoter	/translation="MSGITPRPQQTTTDLFSALSGKSPQNTFGEONTOQAIDPSA
gene	LLFGSDPTOKDNVFGTSDTVQNPQADSKPNDQSNTAKLISALISLQMLTNSNKKO
promoter	DTNQEDQSOAPQNNNGGLGTPSADSGGGTPTDGTGGGGTPTATGGGGGTPTATG
gene	GGSGGGGTPTATGGGGTPTATGGGGGTPTATGGGGGTPTATGGGGGTPTATGGGGGTPTATG
promoter	ENKGIENVKDTIKVGAQVDFDGHGATFDADKSMGNDGKQKPFELAGATLANK
gene	LDNGENVDGDIHYKAKNAQEVTDINHAQNVGDLITVKGEGGAAYNLNLIKNSSSAKGA
promoter	DDVVLQNLNTHLKIIDNFKADDFGTVMTNGGKQFDMSIELNGIEANHGKFKALVKS
gene	SDDKLATGNIAMTIDKHYADKDTQASTQHTL"

[illegible]


```
FT FT PWTDEAVANATRFRRSPDLSHQSDGTRISMLATDPDQSPSSGSKIGDSGPIPPREP
FT FT MLWNSNGRFFELQDEKLVNRSEPOGSIQDLAKRPDPFSTENPTGLAPLDDLSILATPKQT
FT FT YLHQSKDQVGHGHOQLQAQHFHLLAQLDSSLAIVRSNEALLIEGKPPKPAVKMEREDG
FT FT NIHDTASGRKTQPLGKAHIAHTJTNVLSSHGGRMRVHEDRLYQFDPISPRWKIPGL
FT FT EDAFNSLSITGNSVYAKSDADVDDVLLSSHPMHVEYEDLOSFSVADPNRAALLSGKTT
FT FT QAILLTDHSPYIGLTPKTKGLBELDGGKAQAAVGLSGDKLFTADTQGRUISADRSAF
FT FT EGGDKPLKMPENQANFOLEGGVPLGGHNRKVTGFINGGDGGVHALIKNROGETHSHALDEQ
FT FT SKSQSGWNLNVALVNNRGLTTPPPPTAADRLNDRAGLVLSGEGIQRWMDATPCW
FT FT KEAGIKDRLQAGDSNAYVLKGGKHLAKIAAEHPNMAFDRNTALAKSPKVBEMG
FT FT DEIGLDDRVTKAFAMYSNREVALDDQNKLTAAHSKDHKPVTLDPGLEGGDKISLSIDE
FT FT KHNHLATSTGLGLCLPKEAWQSKLGDGLRARTPTVALPGGQPVKALFTNDDNVLSAQ
FT FT IEDAEKGLMQLKQWQRFQRPVNTAEPTPMKNVGDHIQHRYOGRGLKVEYETESM
FT FT FGRSGVEKSKASTSEFIRANYKTAETPRMKNVGDHIQHRYOGRGLKVEYETESM
FT FT LFKQLEIHESGGRPPARQDGLKARIATALEAKLPGQGLVKELETLRDELENHSYAL
FT FT MSIQGSGYKAKNLKQDQILNQHGLAKSPVMOFGKGLADLGTKLPKSSGHDLVKEL
FT FT QDALTOVAPSANPTKLLGTGLKHQGLKLSHOKADIPLGQRORDASDHGSKARLALDL
FT FT VTLKSGALLDQVQLPOSDIEPLQKGLATLRDVTYGENPVKVYEDMGPTDNKALESG
FT FT YESVYKTLKSKADHAVSNMRAATSGKQDAELAGFKSKMLKOLEHGGDDRVGLQRSYG
FT FT VNLTPFIIADKATGLWPTAGATGNRRNYILNAEREGGVLYLISSEGAVNSGGFGAG
FT FT KDWPGFDIANDNARSVDNNRNLTPNFRGLVDVATVAASQAGVNVFVNPDESDIDAF
FT FT VDLLEFQOLNPQLKKAHDHESYEAREFNEDTAGGTADIRAGINLTEDRDPNADPS
FT FT DSFAVVRGGPAAANTVNLMTYTOYSLTKNDKTELKGGKRRRFLNNVTAGGOLRAQ
FT FT IGGSHPTPTGPASAPGPTPASOTAAANLGGALNFSVENRTVKRIKRYNVAKPITTEG
FT FT LSKLSKGLGEAPLNTKAKLAELADPLNARYTGCKKDEVIQAQDLGLEELFADIPPPK
FT FT DNDKOYKALRODRAVKAHRASANKHVSMDNAREFTSKTNLSGUSSESLTKINSSVRD
FT FT ASAPGNATRVAEFRQDPKRLKAMKEMEGSIGTLARVLRPEKDSLVKDIKDEGSLUNGMT
FT FT QDLSLMELEDNEMRIKRLVVFHTATQAEENFTSPTPLVSYNSGANVSVTTLGRINEVY
FT FT GADQDKPIGYTFDGELSRPSASLKEAAGDLKKEGFELKS"
FT FT complement(8226..8621)
FT FT /codon_start=1
FT FT /db_xref="SPTREMBL:Q9JP37"
FT FT /note="ORF2"
FT FT /transl_table=11
FT FT /product="unknown"
FT FT /protein_id="AAF71500.1"
FT FT /translation="MTWKTSPDPAFTINSIGAOLGTLTLONGVCALYDQONNEAII
FT FT ELPHSEMVIFHCIRGCRPERAPDLRLLSLNFVARLHGCWFVADQDGVRLCAQRELA
FT FT SLDPAPCDVTRGTFSQAREARFLQA"
FT FT complement(8960..10831)
FT FT /codon_start=1
FT FT /db_xref="SPTREMBL:Q9JP36"
FT FT /note="ORF3"
FT FT /transl_table=11
FT FT /product="unknown"
FT FT /protein_id="AAF71501.1"
FT FT /translation="MISRIIGGAGGKLSRVNQOHDVPAQTAHPNNAVTAAGMNPPLTPD
FT FT QSGSHAYTESSAGARLNVAARHTQLLQAFKAHEGTAPVSGAPMISRAALLIGSLIQA
FT FT EPLPEYMAEKLSPERVQLKQFQSDILOQRLEKFAQPGQIPDKAEVQLIKGPAQSVAD
FT FT QLEHFQMLHDAEPATVGOHAKADKATLAVSQTAGYAGRASKAIGBGLSNSIASLDEH
FT FT ISALDLTLDHDAEQNKESLHADRLQALVDKTLVLGHADFYKSPKARLXASVAAHTQID
FT FT NVVSDLVTRNTVGGWKAGP IVAAAVPOQLFSSMTHLGYRLSTSDKLRDITIPETSDA
FT FT NMLKASIGIVMAGIAHETVNSVVKPMFQAALOKTGLNRLNLMVPMKAVDNTVTPDPFE
FT FT LKSHGELVKTPEEVAODKAFVKSERALLNKVKQGSSTHPVGEIMAYSAFGCSQAVR
FT FT QMLNDVQINGQTLASALASGFCGAVSASQTLQLKSNVDPQGRKKIPVFPDRAES
FT FT DMLKDLKGMDLRFPSTVRTYTSKALSGIOSSALTPPVTAQAEAGASGTLISAGATLR
FT FT NMLAATGSGSLTYTNSVTVAEAKLKAAGMGATPMLDRTEIL"
FT FT complement(110828..11322)
FT FT /codon_start=1
FT FT /db_xref="SPTREMBL:Q9JP35"
FT FT /note="ORF4"
FT FT /transl_table=11
FT FT /product="unknown"
FT FT /protein_id="AAF71502.1"
FT FT /translation="MTNNDQVHTLI NEICALSLISTPERFVESAENFKISEVDFTLOFQD
FT FT RDEGRAVLIVGDMCALPARGESALLALMDINFMHFAGASHAPSFNAQTCRVLMLGCV
FT FT ALERASRAGVLLLMKSFSDIAKKEWHGFMGQATAGSSSTDQPVAPAAKRSLSAPORF
FT FT Q"
FT FT 11628..12902
FT FT /codon_start=1
FT FT /db_xref="SPTREMBL:087264"
FT FT /transl_table=11
FT FT /gene="hrpw"
FT FT /product="hrpw"
```

```
FT FT /protein_id="AAF71503.1"
FT FT /translation="MSIGITPRPOQTTPPLDFLSALSGSKSPQNTFGSQNTQQAIDPSAL
FT FT LFGSDTKQDNVFGTPDSTVONPODASPNDSQSIKALISALIMSLQMLTNSNKKQDT
FT FT NOEQPDSQAPOPNNGGIGLTPSADSGGGGTPTATGGGGGTPTSGGGGDDTPTATGGG
FT FT SGGGCTPTATGGGSGGTPATATGGGEGVTPQITPOLANPRTSGTSYSVDTAGSTEAG
FT FT KINVKQDITKVGGEVDFGHGATTAADKSMNGDQGNQKPMFELASGATLKNNVLNG
FT FT EVGDIHVKANQAQEVITDINHAQNVGDELITVKGEGAAVTNLRNKSASAGADDKVLA
FT FT LNAATHLKIINFRADEFTVTRTNGGQFDDMSIELNGIEANHGKFAVKRSDDSLKLA
FT FT TGNFAMTDVHAHYDKTQASTQHTEL"
FT FT complement(12986..14446)
FT FT /codon_start=1
FT FT /db_xref="SPTREMBL:Q9JP34"
FT FT /note="ORF5"
FT FT /transl_table=11
FT FT /product="unknown"
FT FT /protein_id="AAF71504.1"
FT FT /translation="MHNRRVQOPVTTATDSFTASDASLASSSVRSVSDQOREINAI
FT FT ADYLTHVFAHKLPPADSADQANVDVHNAQITALTIETRSLRHIFGEPATPATDTPA
FT FT KAEKLDRLMTTSGALRATPFAMASLIQYQWPAINKGDWLPAPLKPRTPLISGALSAM
FT FT DQVTKMMDRATGDLHYLSASPDRLHDMAMASVKRHSPLARVLDTGVAVQYTSARNA
FT FT VRTVLPALASRAVAVQAVDLGVSMAGLAANAGFGNRLLSVQSDRHQRGALVGLKD
FT FT KEPKAQISENDWLEAVKAIKSASYSCAALNAGKRMAGLPLDMATDAMGAVRSLSVSS
FT FT LTONGIALAGSFAGVGKLOEMATKNIITDPATKAVSOLTNLAGSAAVFAAGTTAALTJD
FT FT PAKKASRFDQDTPVKSSTSTTGVVADQTVKLATVTKDMGGEATHTTGASLRWTVNLR
FT FT QVRKAREADIEEGGTAAASPSEIIPFPMRS"
FT FT 14818..15891
FT FT /codon_start=1
FT FT /db_xref="SPTREMBL:Q9JP33"
FT FT /note="ORF6"
FT FT /transl_table=11
FT FT /product="unknown"
FT FT /protein_id="AAF71505.1"

Query Match 11.5%; Score 155.2; DB 16; Length 20052;
Best Local Similarity 57.18; Pred. No. 6.4e-29;
Matches 330; Conservative 0; Mismatches 233; Indels 15; Gaps 2;

Qy 745 ggcgctaatacagcggctgctgacaccattaccgtgaaagcgggtcaggtgtttgat 804
Db 12288 GCCGCAAGATCAATGTGTTGAAAGACACACCATCAAGTCGGCGTGGCGAAGTCTTTCAC 12347

Qy 805 ggcgaagagaaacacctcaacgcgggttcagaattagcgatggcgcgcagctcgaagaac 864
Db 12348 GCCACGGCGCAACCTTCACATGCCGACAAATCTATGGGTAAACGACGACGCGGCAAAAT 12407

Qy 865 cagaaaccgctgttatactaggagacgggtgccagcctgaaaaacgtcacctcagtgccac 924
Db 12408 CAGAAAGCCCATGTTGCGACCTGGCTGAAGCGCTACGTTGAAGAAATGTGAACCTGGGTGAG 12467

Qy 925 gacggggcggtggtattcattcttttacg-----gtgatgcaaaaatagacaaat 972
Db 12468 AACGAGGTGCGATGTCATCCACGTGAAGCCAAACCAACGCTCAGGAAGTCAACCATTTGACAAAC 12527

Qy 973 ctgcacgtcaccaacacgtgggtgagagacgattaccgttaagccaaacagcgcgggcaaa 1032
Db 12528 GTGCATGCCCAAGAACGTCGGTGAAGACCTGTATTACGGTCAAAAGCGGAGGCGGACGCG 12587

Qy 1033 aaatccacgcttgaatacaactaacagttcccttcagcacgcctctgacaagaatcctgcag 1092
Db 12588 GTCATAATCTGAACATCAAGAACAGACAGTGCCTCAAGAGTGTCAGACGACGAGTTGTCGAG 12647

Qy 1093 ctgaatgccgatactaaactgagcgttgacaacgctgaagccaaagacttttggtactttt 1152
Db 12648 CTCAAAGCGCAACACTCACTTCAAAATCGACAACATTCAAAGCGCGACGATTTCGCGCACGATG 12707

Qy 1153 gtacgcactaacgcgggtcacaag---ggtaactgggatactgaaatctgagccataatcagc 1209
Db 12708 GTTCGCACCAACGGTGGCAAGCAGTTTGTATGACATGAGCAGTCGAGCTCGAAGCGGATCGAA 12767

Qy 1210 gcagaagacggttaagtctctcgtttaaaagcgatagcgagggcggttaaacgctcaatacc 1269
Db 12768 GCTAACACGCGCAAGTTCGCCCTCGTGTGAAAACGACACGTGACGATCTGACGCTGCGCAACG 12827
```



```

/protein_id="AAC67261.1"
/db_xref="GI:2618832"
/translation="MKSLSLIMIGFALSLLAAGSNDAKEETKDTGSKTEATASEG
BELVQOSVCGHGRDLEGVSGNQLQEVGKDEHKIESIKNRGRNMPKGLVDDNEAA
VIAWLSKK"
RBS
2806..2813
/gene="ftsE"
2806..3502
/gene="ftsE"
2816..3502
/gene="ftsE"
/codon_start=1
/transl_table=11
/product="cell division ATP-binding protein"
/protein_id="AAC67262.1"
/db_xref="GI:2618833"
/translation="MIENKVEYKAYPNCVKALNGISVTIHRGEPVYVVGSGACKSTF
IKMIYREKPTKGOILNHNKDLATIKKEIIPFVRKRGIVVDFDKLPKLVFENAF
ALEVIGEPSVKKRVLEVDLVLQHKARQFPDQSGEQQRYSIARSIVNPDVVI
ADEPTGNLDPTDSWEVMTLEEINNRGTTVVMATHNKIEIVNTMKRRIAIEDGIIVR
ESRGEYGD"
3476..4385
/gene="ftsX"
3476..4384
/gene="ftsX"
3495..4385
/gene="ftsX"
/codon_start=1
/transl_table=11
/product="cell division protein"
/protein_id="AAC67264.1"
/db_xref="GI:2618835"
/translation="MIKILGRHLRESFKSLGRNTWTFASIAVTVLILVGVFLVIM
LNLNMTAENKQVEIKVLIDLADOKAQLNDIKELKGIQSVTSSKKELDOLV
DSFGDSCKSLTMKQENPLNDAFVVKTDPTDPTNPAKKIEKMDHVYKVTYKGEYSR
LFVKVGRNIGIALIIGLVFTAMFLISNTIKITIFARRKEIEIMKLVGATNWRIRWP
FFLEGLDGVFGSVIPIALVLTSTYQYVIGVWVPKVGQSFVSLLPYNPFVQVSLVIA
IGAVIGWVGSLSIRKFLRV"
4511..4516
/gene="yvjB"
4511..5967
/gene="yvjB"
4525..5967
/gene="yvjB"
/codon_start=1
/transl_table=11
/product="putative protease"
/protein_id="AAC67263.1"
/db_xref="GI:2618834"
/translation="MNQKIMAVIAAGSMLFGAGVYAGINLEMDKPTQAAVPATAQA
DSERDKAMDTEKAYELISNEYVEKVDREKLEGAIOGMLSTLNDPYSVYMDKQAKQ
FSDLSDSFEGIGAEVGMEDCKIIVSPFKKPAEKAGLKPNDIEIISINGESMAGKDL
NHVLLKTRGKGSVSMKIORPTKKOLSPRIKRAEIPLETVEASEKKVQGHVGYIA
ISTSEHTAEDPAKALREKKEIEGLVIDVRGNPGGYLQSVBEILKHFVTKDQPYIQ
IAERNGDKRTFSLTHKKATPVNVTIDKGSASEILAGALKAEAGHYDVVGDTSFGK
GTQVQVPMGDSFNKILTYLKWLPNGNWIHKGIEPTIAIKQPDYFSAGPLQKEPL
KVLNNEDVKHAQVLLKGLSPDGRDGYFSKMKKAMVAFQDNKLNKTKTGVIDTRTA
6773..7977
/gene="yvjD"
6773..6777
/gene="yvjD"
6784..7977
/gene="yvjD"
/note="transmembrane protein"
/codon_start=1
/transl_table=11
/product="yvjD"
/protein_id="AAC67265.1"
/db_xref="GI:2618836"
/translation="MSVQMGTELLKSAGLFFLHPFWFFIITLAFGVYRIKREKTF
HTRADIYDCLKFTYTKGLIGLGLLSVILGGLGISPLGLAIATVITAAAFILRAN
WMSAAYIVSVSMLIGFLGIQYQAEFFLERPFQGFVAVWPAVAVFLGLLIITTEGAVAYR

```

```

SAHVRTSPALVSSRGLPQOOLANRVMLLPLFLVPGNGLESLSHSMWPVFTVPGGSF
HFLATPYFVGFGORVQGSPLPETSIRITAKRVCIILGLAVAVLGAASLLWTPLAGAAVCT
ALLNGFELSIKORUNDNAAPYFYSKRDOGLVLCIIPNTPAEDLELKGEITTKYNGI
PVKNVSDYEALQHNRAIVTKLEIILGNGEIRFQDRAGEGHEHGLILFVKDDREDEA
VASGS"
complement(8025..9373)
/gene="yvka"
complement(8025..9359)
/gene="yvka"
/note="belongs to the major facilitator family"
/codon_start=1
/transl_table=11
/product="yvka"
/db_xref="GI:2618837"
/translation="MSSRKWALIVLLGAILVPINSTMIAVALSSISHTYNBSIAS
ITWVTVLIYMAVTQIATAGLGMNKYKMTLMGVLGLFLIASLGCALSPSLLLIIVF
RALQAVGALLTPNSIAIRHVSEKRLPKVFGFGLGAGLGAALPGFISLIDISFS
WHSIFWNPFLAIALFTALTMFPQYKENSADLDITGSLLAGSIVSILITKNEA
PWGTVTVSVLILFPLFRREKTOHPIDFALPKSFTTNANLSVLSLMMYAVL
LIMPLEWNOFGLATNSGMALSVFSIFMSASNMVGAOLHHKWKAKKILIFLSEFAMAG
ANLLFLILSSSHSVLEFLMSLILGLGASVGLTSMQVSSLATVDPGMSGVASGIFSTF
RYFSIISALIGLISGTHLFLMILFAVSIILGVFVSIKSDETARIEKNSA"
complement(9366..9373)
/gene="yvka"
complement(9390..9972)
/gene="yvkb"
/note="9959"
complement(9390..9959)
/gene="yvkb"
/note="belongs to the TetR/AcrR family of transcriptional
regulators"
/transl_table=11
/product="yvkb"
Query Match 6.8%; Score 91; DB 1; Length 47739;
Best Local Similarity 58.8%; Pred. No. 1.4e-12;
Matches 157; Conservative 0; Mismatches 110; Indels 0; Gaps 0;
QY 748 gctaatacagagtgctgcatacaccattaccgtgaaagcgggtcagggtgtttgatgac 807
|| || || || || || || || || || || || || || || || || || || || || ||
Db 37642 GCAGCTGACAAAGTGTGTGCACGAAACAAATTATCGTACCAAAAAATACAAATATGACGGG 37583
QY 808 aaagacaaacaccctcaccgcgggttcagattagcgatgagcgccagtcctgaacacacg 867
|| || || || || || || || || || || || || || || || || || || || || ||
Db 37582 AAAGACACCGGTTTGTGGCAGGGAAGAAATTTAGGTGACGGAAGCCAGTCAACAACCA 37523
QY 868 aaacgcgtgttatactggaagacgggtgcagccatgcaaaacgcctcaccatggcgacgac 927
|| || || || || || || || || || || || || || || || || || || || || ||
Db 37522 GAQCCCTGTTTTCGTGTGGAGGATGGAGCAACCCCTGAAAAATGTGTGCTTGTGCACCT 37463
QY 928 ggggagcgatggtatcatctttacagtgatgcaaaatagacaatctgcacgtcaccaac 987
|| || || || || || || || || || || || || || || || || || || || || ||
Db 37462 GCAGCTGATGGCGTGCACACTTATGGAACGTTAACTTACGAATGTGAATGGGAAGAT 37403
QY 988 gtgggtgagagcgcatcaccgttaag 1014
|| || || || || || || || || || || || || || || || || || || || || ||
Db 37402 GTTGTGAGGATGCGCTTAACGCTGAAG 37376
RESULT 11
BSUB0018 BSUB0018 209510 bp DNA BCT 26-NOV-1997
LOCUS Bacillus subtilis complete genome (section 18 of 21): from 3399551
DEFINITION to 3609060.
ACCESSION 299121 AL009126
VERSION 299121.1 GI:2635827
KEYWORDS
SOURCE Bacillus subtilis.
ORGANISM Bacillus subtilis.
Bacteria; Firmicutes; Bacillus/Clostridium group;
Bacillus/Staphylococcus group; Bacillus.

```

REFERENCE
AUTHORS

1 (bases 1 to 209510)
Kunst, F., Ogasawara, N., Moszer, I., Albertini, A.M., Alloni, G.,
Azevedo, V., Bertoro, M.G., Bessieres, P., Bolotin, A., Borchert, S.,
Boriss, R., Boursier, L., Brans, A., Braun, M., Brignell, S.C.,
Bron, S., Brouillet, S., Bruschi, C.V., Caldwell, B., Capuano, V.,
Carter, N.M., Choi, S.K., Codani, J.J., Connerton, I.F., Cummings, N.J.,
Daniel, R.A., Deniot, F., Devine, K.M., Dusterhoft, A., Ehrlich, S.D.,
Emmerson, P.T., Entian, K.D., Errington, J., Fabret, C., Ferrari, E.,
Fauler, D., Fritz, C., Fujita, M., Fujita, Y., Fuma, S., Galizzi, A.,
Galleron, N., Ghim, S.Y., Glaser, P., Goffeau, A., Gollightly, E.J.,
Grandi, G., Guisepi, G., Guy, B.J., Haga, K., Haiech, J., Harwood, C.R.,
Henaut, A., Hilbert, H., Holsappel, S., Hosono, S., Hullio, M.F.,
Ikeya, M., Jones, L., Joris, B., Karamata, D., Kasahara, Y.,
Klaer-Bianchard, M., Klein, C., Kobayashi, Y., Koetter, P.,
Koningstein, G., Krogh, S., Kumano, M., Kurita, K., Lapidus, A.,
Lardinols, S., Lauber, J., Lazarevic, V., Lee, S.M., Levine, A., Liu, H.,
Masuda, S., Mauel, C., Medigue, C., Medina, N., Mellado, R.P.,
Mizuno, M., Moesti, D., Nakai, S., Noback, M., Noone, D., O'Reilly, M.,
Ogawa, K., Ogilwara, A., Oudega, B., Park, S.H., Parro, V., Pohl, T.M.,
Portetle, D., Porwollik, S., Prescott, A.M., Presecan, E., Pujic, P.,
Purnelle, B., Rapoport, G., Rey, M., Reynolds, S., Rieger, M.,
Rivolta, C., Rocha, E., Roche, B., Rose, M., Sadaie, Y., Sato, T.,
Scanlan, E., Schleich, S., Schroeter, R., Scoffone, F., Sekiguchi, J.,
Sekowska, A., Seror, S.J., Serror, P., Shin, B.S., Soldo, B.,
Sorokin, A., Taccioni, E., Takagi, T., Takahashi, H., Takemaru, K.,
Takeuchi, M., Tamakoshi, A., Tanaka, T., Terpstra, P., Tognoni, A.,
Tosato, V., Uchiyama, S., Vandenbol, M., Vannier, F., Vassarotti, A.,
Viari, A., Wambutt, R., Wedler, E., Wedler, H., Weitzenecker, T.,
Winters, P., Wipat, A., Yamamoto, H., Yamane, K., Yasumoto, K., Yata, K.,
Yoshida, K., Yoshikawa, H., Zumstein, E., Yoshikawa, H. and
Danchin, A.

TITLE
The complete genome sequence of the gram-positive bacterium

Bacillus subtilis
Nature 390 (6657), 249-256 (1997)

98044033

REFERENCE
AUTHORS

2 (bases 1 to 209510)
Kunst, F., Ogasawara, N., Yoshikawa, H. and Danchin, A.
Submitted
Direct Submission
Regulation de l'Expression Genetique, 28 rue du Docteur Roux, 75724
Paris Cedex 15, FRANCE. E-mail: moszer@pasteur.fr,
adanchin@pasteur.fr Phone: +33 (0)1 45 68 84 41, Fax: +33 (0)1 45
68 89 48

FEATURES

source Location/Qualifiers

1. .209510
/organism="Bacillus subtilis"

/strain="168"

/db_xref="taxon:1423"

complement(4. .626)

/gene="yvyqK"

complement(4. .29)

/gene="yvyqK"

complement(45. .626)

/gene="yvyqK"

/function="unknown"

/note="similar to hypothetical proteins"

/codon_start=1

/transl_table=11

/protein_id="CAB15320.1"

/db_xref="GI:2635828"

/db_xref="SPTREMBL:O34899"

/translation="MKLYTKGDKGTGLVGGRTDKDSLVRVESYGTIDELNSFIGLAL
AELSGOGFEDLTAEILLTQHLEFGDGLAIIVTERKDYKLFTEESYFLTRIDAVTA
EAPELKKFLPGSKSCASLLHARTITRRAERVAALMKEEITHEIVLRYLNRLSDYF
FACARVYNARSGIGDVEYERSAIVFRDRNSSSES"

complement(649. .1977)

/gene="yvyrA"

complement(649. .1977)

/gene="yvyrA"

/function="unknown"

/note="similar to iron transport system"

/codon_start=1

/transl_table=11

gene

/protein_id="CAB15321.1"
/db_xref="GI:2635829"
/db_xref="SPTREMBL:O34631"
/translation="MKAREGLSGGDSRLNNVSLVVEKGEFLGILCPNGSGKTTLLH
LTGTLTPAKKGRVYLAKLADYKPKELAMVLPKMQQQAETFTVEETVAGRTYF
QTGLFQROTKEGEAIVOEAMEQTGVADFAQKPIRELSGGGQQRVYLAQALAQPRILF
LDPTNFDLAYQKDLDLIKRLTRESGLAAVDFHDLNTASLYCDGLMFMKNGTAGP
KQPEYAVTEQSIKAVYDVTALVHQSSPKMIVIOPEKDSVKRQSIPEFALLQAGR
DDLLOLEPLRTLSSTPICAGFSWSTRLTHKRLPDQDPDIEGLTACLSSGQLOQT
CMASSERLDRFVRYVEDGELSVEFCVQTFGSWILLINGCYAADOFFIKALMAEAEER
TKVLGGGGTGDILIAATQQTQSENIEQRNLQIKKGTAECIKAEAELEF"

gene

complement(1977. .3038)
/gene="yvrB"
complement(1977. .3038)
/gene="yvrB"
/function="unknown"
/note="similar to iron permease"
/codon_start=1
/transl_table=11
/protein_id="CAB15322.1"
/db_xref="GI:2635830"

CDS

/db_xref="SPTREMBL:O34451"
/translation="MKKAFTRIRLRNNRYAAYIGFSFLAVSLIGISCSGLHIPIPA
VFRVHQGGSGSDPMYTNIMNIRLPVYLAALVGAALSIAGAATFOGLKNPL
ADPTLGVSSGASVGAVTTLGLHLFPVIGGFTLPVLSVAALATMAAATVFFSRFLHA
SMVSSTLITGVITNSPLGAFISLIITALTGNLLPIVHMLLGSVMRGSVYILFLPF
FLGTLTLLINGRELNVMTYGEDKAKILGVSVQORKMWILLIAGSLLTGSAVAVSGTIG
FVGLVPHITRLNGDTHRHLLPLSALLGAGFLVADLLSRTIIEPIELPIGIIITSLA
GAPVALILIRHQHGRSL"

gene

complement(3001. .3945)
/gene="yvrC"
complement(3001. .3945)
/gene="yvrC"
/function="unknown"
/note="similar to iron-binding protein"
/codon_start=1
/transl_table=11
/protein_id="CAB15323.1"
/db_xref="GI:2635831"

CDS

/db_xref="SPTREMBL:O34805"
/translation="MKKRAIGIAWALLLAAYMLAGCGNPADQKSKAKQKTEVPVITD
DASNDQVTIKKEPKIVSLSPNTEITAYLGLGDKVGVVTTNDTYPEVKKVEKVGDM
NVNVEKVISLPLDLVAHESMSASADAIKQLDAGITVLTVNDAAQSFSEVYSIEMI
GEAGAEKADQLVKSHKSDLKIOEKAKTISKDEKSVFIEYSPDPDITVTKGDTFM
KHMNVTHAKNAADOTGWQMTDEATVKNLPPDAIVTDCGKAKAVEKRDGWEINAV
KHRVYDVPDLVTRSGPRLIEGVEELAESIYDPDFKE"

gene

4343. .5165
/gene="yvrD"
4343. .5134
/gene="yvrD"
/function="unknown"
/note="similar to ketocacyl-carrier protein reductase"
/codon_start=1
/transl_table=11
/protein_id="CAB15324.1"

CDS

/db_xref="GI:2635832"
/db_xref="SPTREMBL:O34782"
/translation="MDLQREKLVLTGTSSTGIGKAAKASFLOEGAAYVNGRKQETV
DRTIELSGYGVHGAADLSKDEAAAFIEKVEIGDIDILVNNLGGFFVEKDFADVT
DEENQYFEYNVMSAVTRSHFLPKMLAKNSGRILNIAESAGVKPLPTMIPYSMTKTA
LISLRGMAEMTKGTNTVNSVLPGPPTWTEGVASVMEGAQAAGQDPTDTIKDYKVN
EPTSLTORYATAEAVANTIVFLASDAASAINGTAQRVEGGIIRSL"

terminator

gene

complement(5172. .6050)
/gene="yvrE"
complement(5172. .6050)
/gene="yvrE"

CDS

/function="unknown"
/note="similar to RNA polymerase"
/codon_start=1
/transl_table=11
/protein_id="CAB15325.1"

```

/db_xref="GI:2635833"
/translaton="MDAVALRADTRAVIGEGPLWDEENGRLYWVDILGSELHIFDPEEK
KNSIPKSFVTALAKYSKDELATMTKMDGLFYLYHLRDSLEKI QPKDMHESIRNFMDA
ICDYPGLRWAGTTSMEGEQKOASLYRLNLDSGLVKTKDOVTSNGLDWDRERLWYVI
DPTQETVRYSDYDQSDNSPEPVYRFDOSDGLPGDQMTIDQNGMLWVALFGLGSRVHV
IDPQKEINSISVPAYKYTCCAFGGRDLLTLITYTATEQMTKEKRYEQPHAGLGLFSA
QLETGGYVPVFFAGDV"
gene
complement(6122. .7843)
/gene="yvirg"
CDS
complement(6122. .7843)
/gene="yvirg"
/function="unknown"
/notes="similar to two-component sensor histidine kinase
[yvirH]"
/codon_start=1
/transl_table=11
/protein_id="CAB15326.1"
/db_xref="GI:2635834"
/db_xref="SPTREMBL:O34989"
/translaton="MEHFFGQMLIVLTLLTMLVAFYYLDARFSAESNSGLTKAT
DTLEAYLDVNEDETWEVDNFKLSYDKQHGWMQIIDSSEGTDSYGVDPDQVGTYYKK
ELSTIYIKRKLHKNKAIWNAINEDKSYLLSGWKSKEQLLTSVKREOKIDSLRYS
KSSITDYIKRKGAIYLLDSNGKILDSINSTKERTMNOLELLKYSKRPWNYKREIS
VTLNKDRMWAVPNPVYTDGEFNKSLFVLYKAMFLVAVLFMYIITWTVYMYFR
FGLPIFTIRWLVNLSKGLKEUENREGRPVSKNKKGKIQKPYRFEGEIEFESMDQLTE
TLRDKSNREKIOATREWTAGLSHDLLTSLSYGVSMLESKQDWSPCEEGEMGO
VVRKSYMSKLLDLNLTLYLAKNDALPIERKLTSLIPPKNVIEDPKKNPFSQVDI
SFVSKBEHIEFALDEAWFRILENLGNAYKHNGKGTETQVILEQTKNHLSLKVQNG
KGMDEETIHLFNRYRTNGTNTKOSTAGTGLGLALAKELVHLHNGTITHVNSRTNIGTV
TILFKKO"
gene
complement(7861. .8970)
/gene="yvirH"
CDS
complement(7861. .8970)
/gene="yvirH"
/function="unknown"
/notes="similar to two-component response regulator [YvrG]"
/codon_start=1
/transl_table=11
/protein_id="CAB15327.1"
/db_xref="GI:2635835"
/db_xref="SPTREMBL:P94504"
/translaton="MDGQFQKQKKQKDETYDIEHLIACFSPMIRKKLSNTSYOEREDL
EQELTKMFKAQMLLCQDYPGWFEPTLYMVDENSSIISHMKKHQKALVCSFPAVLVI
Query Match 6.8%; Score 91; DB 2; Length 209510;
Best Local Similarity 58.8%; Pred. No. 1.4e-12;
Matches 157; Conservative 0; Mismatches 110; Indels 0; Gaps 0;
Qy 748 gctaatacagacggtgctgcatgacaccattaccgtgaaagcggtcagggtttgatggc 807
Db 190191 GCAGCTGACAAGTGTGTGCAGCAACAATATATCGTACCAAAAATACAAATATGACGGG 190250
Qy 808 aaagggacaaaccttcacgcgcggttcagaattagcgatggcgccagctgaaacacag 867
Db 190251 AAAGCAGACGCGTTGTGGCAGGGAAGAATTAGGTGACGGAAGCCAGTCAGAAAACCAA 190310
Qy 868 aaacgcgtgtttacttggaagacggttcagcagctgaaaaacgtcaaccatgggcgacgac 927
Db 190311 GACCCTGTTTTCGTGTGGAGGATGGAGCAACCCCTGAAAAATGTGGTGTGGTGCACCT 190370
Qy 928 gggcgcgatggtattcatcttaacggtgatgccaaaaatagacaactctgcagctcaccaac 987
Db 190371 GCAGCTGATGCGTGCACACTTATGGAACCGTTAAACATTTCAGATGTGAATGGGAAGAT 190430
Qy 988 gtgggtgaggacgcgattaccggttaag 1014
Db 190431 GTTGGTGAGGATGCGTTAACGGTGAAG 190457

```

DEFINITION	Bacillus sp. pella gene for pectate lyase, strain BP-23.									
ACCESSION	AJ237980									
VERSION	AJ237980.1 GI:4585731									
KEYWORDS	pectate lyase; pella gene.									
SOURCE	Bacillus sp. BP-23.									
ORGANISM	Bacillus sp. BP-23									
REFERENCE	Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillus.									
AUTHORS	Soriano.M., Blanco.A., Diaz.P. and Pastor.F.I.									
TITLE	An unusual pectate lyase from a Bacillus sp. with high activity on pectin: cloning and characterization									
JOURNAL	Microbiology 146 (Pt 1), 89-95 (2000)									
MEDLINE	20121742									
REFERENCE	2 (bases 1 to 1214)									
AUTHORS	Pastor.F.I.J.									
TITLE	Direct Submission									
JOURNAL	Submitted (12-APR-1999) Pastor F.I.J., Microbiology, Faculty of Biology, University of Barcelona, Av. Diagonal 645, 08028 Barcelona, SPAIN									
FEATURES	Location/Qualifiers									
source	1..1214									
	/organism="Bacillus sp. BP-23"									
	/strain="BP-23"									
	/db_xref="taxon:89769"									
-35_signal	39..44									
gene	/gene="pela"									
	39..763									
-10_signal	62..67									
RBS	/gene="pela"									
	80..85									
CDS	/gene="pela"									
	95..763									
	/gene="pela"									
	/EC_number="4.2.2.2"									
	/function="pectate degradation"									
	/codon_start=1									
	/transl_table=11									
	/product="pectate lyase"									
	/protein_id="CA840884.1"									
	/db_xref="GI:4585732"									
	/db_xref="SP:PREMBL:Q9X622"									
	/translation="MKKMTLLLSAGLVASIFGVMPAAAPTVVNSTIVVPKGTVDVQKGFVFNSTLQDGSQAEKQKPFLEAGATLKNVIIGAPADGVHCYGSNLSHVVWDVGDGDAITLKSSTGVNTITGGAAYKAYDKVFOMNASGTINIKNFRADDIGKLVKRGNGTSYAVNMTLDNSNISNVRDSINRTDSSVSQGIKTNTRYISKVPTLFGFASCKTSQSGNTQI"									
sig_peptide	95..169									
	/gene="pela"									
mat_peptide	170..760									
	/gene="pela"									
	/EC_number="4.2.2.2"									
	/product="pectate lyase"									
BASE COUNT	329 a 272 c 306 g 307 t									
ORIGIN										
Query Match	4.7%; Score 63.2; DB 2; Length 1214;									
Best Local Similarity	51.9%; Pred. No. 2e-05;									
Matches 168; Conservative	0; Mismatches 153; Indels 3; Gaps 1;									
Qy	742	gccggcgctaatacagacggtagctgcatgacacattaccgtgaaagcggtcagggtgttt	801							
Db	161	gccggcggtgcgcacacggctgctgcaattcaacgattcgtttgacggagcgacacacctat	220							
Qy	802	gattggcaagagacaaaccttcaccgcgcgt---tcagaattagcgatggcgccagtcct	858							
Db	221	gatgcagagggaaraacctttgtatgacgaatccctctacattaggtgacgggtcttcacagcg	280							
Qy	859	gaaacaccagaacgcgtgtttattactggaagacgggtgccagcgtgaaaaacgctacacatg	918							
Db	281	gagaatcagaagccctgtattccggcttgcagacaggcgcaactttgaaaaattgtattattt	340							

RESULT	12		
BSP237980			
.LOCUS	BSP237980	1214 bp	DNA
			BCT
			08-FEB-2000

```

Qy 919 ggcagcagcggcggtgattcattcattcgtgatgcacaaatagacaatctgcac 978
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 341 GGTGACCCGCGGATGGTGCATGTTACGGCAGCTGTATATTTCCAAAGTTGTA 400

Qy 979 gtccaccacgtyggtaggcgcgattaccgttaagccaaacagcgcggcgcaaaaatcc 1038
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 401 TGGGAAGATGTAGCGGAGGATGTCATTCACACTGAAATCATCAGGAACGGTTAATATTACC 460

Qy 1039 cacgttgaatacactaacagttcc 1062
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 461 GCGGAGCAGCGGTATAAGCGCTTAC 484

RESULT 13
FSOPELAB
LOCUS FSOPELAB 884 bp mRNA PLN 27-APR-1993
DEFINITION . Fusarium solani pisi (Nectria hematococca) pectate lyase (pela)
mRNA, complete cds.
ACCESSION M94692
VERSION M94692.1 GI:168157
KEYWORDS pectate lyase; pectate lyase A.
SOURCE Fusarium solani (individual isolate , sub_species pisi) (library:
lambda gt11) population VI cdna to mRNA.
ORGANISM Fusarium solani
Eukaryota; Fungi; Ascomycota; mitosporic Ascomycota; Fusarium.
REFERENCE 1 (sites)
AUTHORS Crawford,M.S and Kolattukudy,P.E.
TITLE Pectate lyase from Fusarium solani f. sp. pisi: purification,
characterization, in vitro translation of the mRNA, and involvement
in pathogenesis
J. Bacteriol. 174, 6343-6349 (1992)
JOURNAL Arch. Biochem. Biophys. 258, 196-205 (1987)
MEDLINE 88022783
REFERENCE 2 (bases 1 to 884)
AUTHORS Gonzalez-Candelas,B. and Kolattukudy,P.E.
TITLE Isolation and analysis of a novel inducible pectate lyase gene from
the phytopathogenic fungus Fusarium solani f. sp. pisi (Nectria
haematococca, mating population VI)
J. Bacteriol. 174, 6343-6349 (1992)
JOURNAL J. Bacteriol. 174, 6343-6349 (1992)
MEDLINE 93015682
FEATURES
    source
        1..884
            Location/Qualifiers
                /organism="Fusarium solani"
                /isolate="T8"
                /sub_species="pisi"
                /db_xref="taxon:108890"
                /sex="population VI"
                /tissue_lib="lambda gt11"
                <1..884
                /gene="pela"
                1..884
                /gene="pela"
                12..740
                /gene="pela"
                /EC_number="4.2.2.2"
                /codon_start=1
                /evidence=experimental
                /product="pectate lyase A"
                /protein_id="AA033339.1"
                /db_xref="GI:168158"
                /translation="MKFTAAVAALVGTSSAAVTKLPKSAGATSFPTAVPVKSGYDG
                WMKRFREPVKQDETKGDKAMFLENGALSNVIGASQAEVHCKGCTCLNNVW
                WADVEDAVTLKQTSYINGGAFHASDKIIQFNGRGTVHVKDEYADYGLKLSRC
                GNCKDNGSPNIVENSVAVDGVLGCGINTNYGDKVINSQDKGKCYCDRYEGNSGG
                KEPTKGGSPDGKICYVTGTTSC"
                12..62
                /gene="pela"
                63..737
                /gene="pela"
                /note="the amino acid sequence of the N-terminus of the
                mature protein and an internal CNBr peptide have been
                determined"
                /product="pectate lyase A"

```

```

polyA_signal 865..870
/gene="pela"
polyA_site 884
/gene="pela"
BASE COUNT 204 a ..242 c 233 g 205 t
ORIGIN
Query Match 4.68; Score 62.2; DB 14; Length 884;
Best Local Similarity 52.0%; Pred. No. 3.5e-05;
Matches 173; Conservative 0; Mismatches 148; Indels 12; Gaps 1;

Qy 854 agtctgaataccagaaacccgctgttatactggaagacgctgagcgcctgaaacagtcac 913
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 193 AGACTGGTGAGAGGATGCCATGTTTCATCTTGAGATGGTGGCCACCTCTCCACAGTTA 252

Qy 914 ccattggcgacgacggggcggtgattcattcattcgttgatgccaaataagacaatc 973
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 253 TCATTGGTGCCTCCAGGCCGAGGGTGTTCACAGGGAACCTTGCACTCTCAACAATG 312

Qy 974 tgcacgtcacaacgctgggtgagcgcgattaccggttaagccaaacagcgcggcgcaaaa 1033
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 313 TCTGGTGGCGGATGTTCTGTGAGGATGCTGCACTCAAGCAGACCGGAAACCTCTT 372

Qy 1034 aatccccgcttgaatacactaacagttccttcgagcgcctctgacaagatcctgcagc 1093
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 373 ACATCAACGGTGGTGGCTTTC-----CAGGCCTCTGACAAGATCATTCACT 420

Qy 1094 tgaatgcgatactaaactgagcttgacaaactggaagggccaaagactttggtactttg 1153
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 421 TCAACGGCGCGGAAACCGTCCACGCTCAAGGACTTCTACGCGGAGACTACGGAAGCTCT 480

Qy 1154 tacgactaacgcggttcacacagggttaactggg 1186
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 481 CCCGAGCTCGGAAACTGCAAGGACAAACGGTG 513

RESULT 14
SC2G38 40549 bp DNA BCT 13-DEC-2000
LOCUS Streptomyces coelicolor cosmid 2638.
DEFINITION AL445503
ACCESSION AL445503.1 GI:10803131
VERSION
KEYWORDS
ABC transporter ATP-binding protein; ABC transporter transmembrane
protein; asnC-family transcriptional regulatory protein; hydrolase;
integral membrane protein; lipoprotein; marR-family transcriptional
regulator; oxidoreductase; molybdopterin binding subunit; regulatory
subunit; secreted deacetylase; secreted lyase; secreted protein;
tetR-family transcriptional regulator; two component system
histidine kinase; two component system response regulator; ung,
uracil DNA glycosylase.
SOURCE Streptomyces coelicolor
ORGANISM Streptomyces coelicolor
Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
1 (bases 1 to 40549)
AUTHORS Redenbach,M., Kieser,H.M., Denapate,D., Eichner,A., Cullum,J.,
Kinashi,H. and Hopwood,D.A.
A set of ordered cosmids and a detailed genetic and physical map
for the 8 Mb Streptomyces coelicolor A3(2) chromosome
Mol. Microbiol. 21 (1), 77-96 (1996)
JOURNAL 97000351
MEDLINE 2 (bases 1 to 40549)
REFERENCE Saunders,D.C. and Harris,D.
AUTHORS Unpublished
JOURNAL
REFERENCE 3 (bases 1 to 40549)
AUTHORS Cerdano,A.M., Parkhill,J., Barrell,B.G. and Rajandream,M.A.
TITLE Direct Submission
JOURNAL Submitted (13-OCT-2000) Streptomyces coelicolor sequencing project,
Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge
CB10 1SA E-mail: barrell@sanger.ac.uk Cosmids supplied by Prof.
David A. Hopwood, [3] John Innes Centre, Norwich Research Park,

```


THIS PAGE BLANK (USPTO)